The Dissertation Committee for Scott Allen Robotham Certifies that this is the approved version of the following dissertation:

De Novo Sequencing and Peptide Characterization via Ultraviolet Photodissociation Mass Spectrometry

Committee:

Jennifer S. Brodbelt, Supervisor

Richard M. Crooks

Andrew Ellington

Jason B. Shear

Lauren Webb

De Novo Sequencing and Peptide Characterization via Ultraviolet Photodissociation Mass Spectrometry

by

Scott Allen Robotham, B.S.; M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2015

Acknowledgements

It is difficult to figure out where to being when saying thank you to those who have played a role in me getting to where I am today, as so many members of my family, friends and mentors have made this possible. Of course, I first would like to thank both of my parents, my mother Kathy Robotham for always supporting me not matter what, and my father Eugene Robotham for instilling in me a sense of personal responsibly and showing me the value of critical thinking in every part of life. I would also like to thank Jade Roth, my favorite cousin, whose conversations went a long way in helping me maintain sanity during graduate school. I would be remised if I did not mention my two closest friends and confidants, Christopher Krotz and Rachel Steadman, who were always able and willing to give me words of encouragement, sharp rebukes, or just a welcome distraction when needed. Also a special thinks to Dr. David Treichel and Dr. Jodi Ryter who refused to let me settle for the easy path and set me on my current course.

I am grateful to my advisor Jennifer Brodbelt, for your support and guidance, for always leading by example and setting the bar high, for always believing in me even in the worst of times – Thanks to your leadership I have gone from a person with a degree in chemistry to a fully capable scientist.

I owe a huge debt of gratitude to the members of the Brodbelt lab, past and present- From training on the proper use of instrumentation or a method; to being a

sounding board for confusing problems and other troubles. I have enjoyed the time spent both in and outside of lab.

Finally, the work in this dissertation would not have been possible without the aid of my collaborators. I would thank Dr. Andrew Ellington and his graduate student Christian Kluwe, for providing interesting samples to analyze. Thanks to Dr. Stephen Trent for growing of cells to be used for sample to optimize new methods. A huge thanks needs to go out to Andrew Horton and Dr. Edward Marcotte whose work in the development of a new algorithm are a corner stone of the work in the dissertation. Lastly I would like to thank Dr. Joe Cannon, Victoria Cotham, and Dustin Holden, all inter group collaborators, whose ideas and skills have played major roles in the completion of this work.

De Novo Sequencing and Peptide Characterization via Ultraviolet Photodissociation Mass Spectrometry

Scott Allen Robotham, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Jennifer S. Brodbelt

Although *in silico* database search methods remain more popular for shotgun proteomics methods, *de novo* sequencing offers the ability to identify proteins lacking sequenced genomes and ones with subtle splice variants or truncations. Ultraviolet photodissociation (UVPD) of peptides derivatized by selective attachment of a chromophore at the N-terminus generated characteristic series of *y* ions. The UVPD spectra of the chromophore-labelled peptides were simplified and thus amenable to *de novo* sequencing.

E.coli lysates were modified by the use of carbamylation and the attachment of a UV chromophore to the N-terminus of digested peptides. UVPD of the resulting peptides generated clean sets of *y* ions. A novel *de novo* algorithm, UVnovo, afforded high confidence identification of thousands of peptides from an *E. coli* lysate and allowed UVPD/CID paired spectra to be searched. *E.coli* lysate peptides were analyzed in alternating scans by UVPD and CID on the same precursors to generate paired

UVPD/CID spectra. UVnovo enabled sequence tag *de novo* sequencing of peptides in order to find matching sequences from a database. Ultimately the UVnovo functioned as a standalone *de novo* sequencing program or a hybrid *de novo*/database search program.

In an effort to better characterize the fragmentation pathways promoted by ultraviolet photoexcitation in comparison to CID, six adrenocorticotropic hormone (ACTH) peptides in a range of charge states were subjected to 266 nm ultraviolet photodissociation (UVPD), 193 nm UVPD, and CID. While both UVPD and CID led to preferential cleavage of the Y-S bond for all ACTH peptides (except ACTH (1-39)), UVPD was far less dependent on charge state and location of basic sites for the production of C-terminal and N-terminal ions.

Table of Contents

Chapter 1

Introduction

1.1 Introduction

Advancements in the field of proteomics over the past several decades are due in large part to the capabilities of mass spectrometry (MS). Improvement of MS instrumentation has allowed investigation of complex cell lysates making it possible to identify thousands of proteins as well as enabling more comprehensive characterization of proteins such as mapping new post translational modifications and sequence mutations. With such endeavors becoming the mainstay of many mass spectrometric-based proteomics applications, more powerful data analysis algorithms have been required in order to support and improve interpretation of MS/MS spectra. Such algorithms are typically trained based on large datasets and optimized to evaluate specific types of MS/MS fragmentation of peptides in order to achieve the best possible confidence. The focus of this dissertation is to develop and optimize tandem MS/MS techniques where peptides are derivatized and fragmented using ultraviolet photodissociation (UVPD) giving uniquely simplified fragmentation patterns. In order to take full advantage of the simplified UVPD mass spectra, a novel data analysis program allowing high confidence identification of proteins and peptides was used.

1.2 Tandem Mass Spectrometry for Bottom-Up Proteomics

For mass spectrometry-based proteomics the most widely strategy is termed bottom-up analysis. In this approach proteins are enzymatically digested to produce more readily analyzed peptides that are representative of the original proteins. The peptides are

separated using high performance liquid chromatography and analyzed using MS/MS.[1] Typically algorithms are used to assign fragment ions of each peptide which is then matched to a protein. There are two different types of programs typically used perform such searches: in silico database search methods and *de novo* sequencing. [8]

In silico database searches take advantage of immense amount of known genomic sequence information by using complex algorithms to match experimental MS/MS patterns to theoretical tandem mass spectra generated in silico for peptides from large databases of known proteins. Since there is no single best way to properly exploit such data, various in silico methods have been developed with the more popular ones being SEQUEST[2], Mascot[3], MassMatrix[4], OMSSA[5], and X!Tandem[6]. Each of these programs provide reliable results while taking different approaches to the processing of MS/MS data to yield peptide and protein identifications. SEQUEST, for example, uses a rapid survey step to identify all reasonable protein sequence based on the MS/MS spectra of peptides, and compares mass to charge ratios for fragments found to the mass to charge ratios of fragment ions predicted for amino acid sequences via a cross-correlation function.[2] Mascot takes into account the intact mass of peptide precursor ions as well as MS/MS data for one or more peptides and applies a probability-based scoring algorithm that calculates a score based on the statistical significance of matching a peptide fragment to fragments calculated from a sequence database.[3] MassMatrix also takes advantage of probability-based scoring for peptides but uses a unique approach by applying two independent scoring models, a descriptive model and a statistical model, where each score is useful for validating a peptide match.[7] Alternatively, OMSSA focuses more on

improving the rate at which data is analyzed by applying a probability-based algorithm and streamlining the analysis.[5] X!Tandem is similarly aimed to decrease analysis time by using a rapid survey step to identify all protein sequences that are reasonable candidates and uses these as models; [8] adapted to an open source system named Global Proteome Machine for which protein identification data could be freely stored and used.[6] Direct comparisons of in silico programs demonstrate that many programs perform at an acceptable level with comparable results, thus leading to ongoing use of all of these methods .[4,9]

The use of mass spectrometry and in silico algorithms to identify thousands of proteins in cell lysate have accelerated many advances in the field of cellular proteomics.[10] Despite the success of this approach, in silico algorithms are limited in that only proteins from organisms with sequenced genomes can be identified. Furthermore, in silico programs have also been shown to suffer in the identification of sequence mutations, truncated proteins, and post-translational modifications (PTM).[11] In an effort to overcome these deficiencies *de novo* sequencing algorithms were developed.[12–14] *De novo* sequencing methods do not require prior information from a protein database; instead tandem MS data is directly interpreted based on the individual masses of the peptides and fragment ions and the mass differences of the fragment ions. By not requiring a fully sequenced genome in order to identify peptides, *de novo* sequencing allows greater flexibility in identifying unexpected mutations or PTMs. Similar to database search algorithms, the complexity required for effective de novo algorithms has resulted in various alternative approaches aiming toward the same goal. The most

common and widely used programs today are PEAKS[15], PepNovo[16], NovoHMM[17], MSnovo[18], and Vonode[19]. Each of these programs utilizes a different approach for scoring the MS/MS spectra, and some of the key scoring metrics are based on predicting the number of amino acids based on observed precursor mass. Scoring metrics are also based on the algorithm making assessments of fragments to determine which peak is selected as the next loss of a amino acid to give a logical path through a MS/MS spectra to construct a peptide sequence that is best fit for the observed fragment ions. PEAKS uses a four step process which involves preprocessing (noise filtering, peak centering and deconvolution), candidate computation, refined scoring, and global and positional confidence scoring.[15] PepNovo focuses on the chemical and physical rules of peptide fragmentation to generate a scoring program.[16] NovoHMM uses a generative hidden Markov model in order to better distinguish between signal and noise peaks.[17] MSnovo uses a probabilistic scoring function with a mass array-based dynamic programming algorithm.[18] The most recent addition to the list of de novo programs is Vonode which uses a unique tag scoring and novel spectral graph to generate *de novo* scoring results.[19] Regardless of which algorithm is used, the success of *de novo* sequencing hinges on the quality of the MS/MS spectra collected. An overabundance of fragment ions from redundant backbone cleavages or neutral losses will inevitably lead to a spectrum that is difficult to interpret. Too few fragment ions (a result of missed cleavages along the backbone of the peptide), however, will result in derivation of only partial sequence information.

1.2.1 Protein/Peptide Product Ion Nomenclature

Dissociation of protein/peptides ions promoted by a given activation method will result in predictable fragmentation often observed as cleavage along the peptide backbone. It is this predictability of peptide fragmentation that allows algorithms to be utilized for searching MS/MS data. A standardized nomenclature has been established for peptide fragmentation which has been accepted throughout the mass spectrometry community.[20] An example of this nomenclature is illustrated in **Figure 1.1** where a letter of *a*,*b*, or *c* denotes a product ion which contains the N-terminus of the peptide and *x*, *y*, or *z* indicates an ion which contains the C-terminus. The formation of *a/x* ions is the N- and C-terminus fragments which are formed by the cleavage of the C_α –C bond, whereas the *b/y* ions result from the cleavage of a N-C amide bond and *c/z* ions are due to the cleavage of a N-C_α bond. A numeric subscript associated with a fragment indicates how many amino acids are found between the cleavage site and the respective terminus. Each pair of fragments is typically associated with certain types of activation methods.

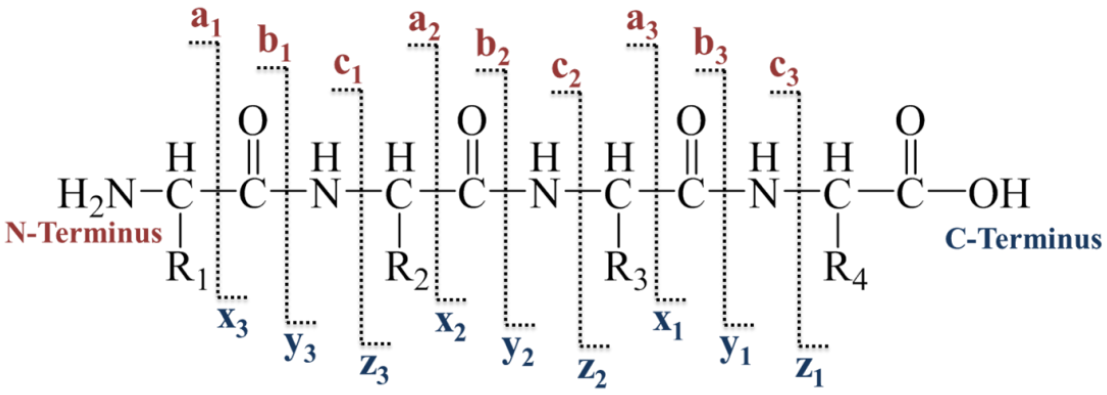


Figure 1.1: Protein/Peptide Fragmentation Nomenclature

For example *b*- and *y*- type ions, which arise from cleavage of the most labile backbone bond (the N-C amide bond) of peptides, are generally the most common and abundant ions formed by thermal activation methods (vibrational excitation). These fragments ions are typically produced by collision induced dissociation (CID) or some photodissociation (PD)[21] methods which have lower energy photons. CID, in particular, is a step-wise vibrational activation process. Peptides undergo radical driven fragmentation upon electron capture dissociation (ECD)[22] or electron transfer dissociation (ETD),[23] which cause cleavage of the N-C_α bond and gives *c*- and *z*-type ions. The *a*- and *x*-type ions are typically only produced from higher energy activation methods such as ultraviolet photodissociation (UVPD)[21] which results in the cleavage of the C_α –C bond of the peptide backbone.

.

1.2.2 Collision-Induced Dissociation (CID)

CID is the most common form of ion activation used today. Ions are fragmented in CID by first isolating a precursor ion and then using a voltage to accelerate the isolated precursor ions into a neutral gas molecules. These collision events cause conversion of the kinetic energy of the ion into internal energy. The amount of energy converted per collision is small, therefore typically several collisions are required in order for a molecule to accumulate enough internal energy for dissociation to occur. CID is currently used in almost every type of mass spectrometer, including quadrupole ion traps (IT),

Fourier transform ion cyclotron resonance (FTICR) ion traps, hybrid orbitraps, tandem quadrupole instruments, and time-of-flight (TOF) mass spectrometers.

In the field of proteomics, the collisional heating results in characteristic backbone cleavage yielding *b*- and *y*- type ions and is the most common method for sequencing peptides and identifying PTMs.[24] Over the years CID has been shown to be an effective and robust method for identifying peptides from complex protein mixtures, so it is no surprise that it is often considered the "gold standard" method of fragmentation. [25] Despite all of the success seen with CID, there are still limitations to its use. Certain amino acids side-chains increase the lability of peptide backbone bonds and with the vibrational excitation of CID preferential cleavage is often observed at these sites (e.g. proline, histidine, glutamic acid, and aspartic acid). In additional to preferential cleavages, uninformative neutral losses such as water or ammonia are common. In addition, labile post-translational modifications may be cleaved, thus preventing site localization of the modifications. Ion traps suffer from an additional disadvantage for CID. As a result of the adjustment of the radiofrequency voltage trapping voltage during ion activation to ensure adequate ion energization, the lower m/z range is truncated.[26] This problem is known as the low mass cutoff (LMCO) and can be detrimental to database search methods looking for low mass PTMs or short N-terminal and C-terminal fragment ions.[27] This limitation has been alleviated by the recent development of higher energy collisional activation dissociation (HCD) in ion traps.[28] HCD is performed by accelerating the ions through a separate collision cell rather than storing them in an ion trap. Fragment ions formed in the HCD cell are then transferred out of the cell for

analysis. By preforming ion activation in this collision cell rather than in the ion trap, HCD is independent of the rf trapping voltage which alleviates the LMCO issues found in CID as well giving higher energy deposition which yields more informative MS/MS spectra.

The general shortcomings of CID have led to the development of alternative activation techniques, such as electron-based methods[22,23], surface induced dissociation (SID)[29], and photodissociation.[21] While it is unlikely that any new method will ever completely replace CID for bottom-up protein analysis due to its well-established usage, other dissociation methods allow a more complete analysis of proteins and are often used in a complementary manner in conjunction with CID.

1.2.3 Electron Capture/Transfer Dissociation (ECD/ETD), Surface Induced Dissociation (SID) and Infrared Multiphoton Dissociation (IRMPD)

Various alternative activation methods have been developed, each offering particular advantages and unique applications. For example SID uses a surface as a target instead of an inert gas. The motivation behind using a surface is thatgreater energy transfer from a single collision with a massive target (the surface).[29] Using SID for peptide fragmentation results in traditional CID-like ion (*b*- and *y*- fragments) and gives comparable sequence coverage to that observed with CID.[30] SID however has also been shown to be an extremely effective tool for probing mechanisms of peptide fragmentation, shedding light on structures and energetics connected with ion formation.[31]

In a rather different approach ECD and ETD are electron-based dissociation methods resulting in cleavage of the N-C_α bond along the peptide backbone yielding *c*- and *z*- type ions. This fragmentation is the result of exothermic reactions between a charged cation and either low-energy electrons (ECD) or radical anions (ETD). These dissociation techniques are radical mediated and have been shown to be effective for characterizing post-translational modifications[23,32–34] and improving peptide sequence coverage.[35]

Finally IRMPD is commonly implemented with a CO_2 laser that emits photons at 10.6 μm. Fragmentation with IRMPD yields in *b*- and *y*- type ions as a result of the peptide absorbing photons via heating of various vibrational modes from C-N, C-C, and C-O bonds. IRMPD is typically performed with a continuous laser during activation which allows secondary dissociation to occur and yielding more diagnostic fragment ions compared to CID.[36] IRMPD also boasts the advantage of having a wider m/z trapping range than CID as rf trapping voltages can be lowered during activation.[37]

1.2.4 Ultraviolet photodissociation (UVPD)

UVPD is an alternative activation method that uses a laser at an ultraviolet wavelength in order to excite peptides based on absorption of high energy photons (from 3.5 eV up to 7.9 eV). While other photodissociation methods using infrared or visible photons have been proven to be effective over the last few years, UVPD has established itself as an extremely powerful technique for analyzing proteins and peptides.[21] The wavelengths used for UVPD have generally been ones readily available from pulsed

excimer or YAG lasers, each taking advantage of a unique chromophore of the peptide or protein for photoabsorption.

157 nm and 193 nm UVPD are unique in that these photons are absorbed by amide bonds, meaning that the entire peptide backbone acts as a chromophore, in addition to aromatic amino acid side chains. Each photon carries a large amount of energy (7.9 and 6.4 eV, respectively), exciting proteins and peptides into higher electronic states. This allows access to diverse fragmentation pathways, including ones with higher activation energies. The resulting product ions from this type of UVPD are often spread across all six fragment (*a,b,c,x,y*, and *z*) types discussed in section 1.2.1. of the Introduction. First reports of such fragmentation date back to the 1980's,[38–40] however, it was not until the past decade that the method started to become more widespread. UVPD using a 157 nm excimer laser has been integrated with both an ion trap mass spectrometer as well as a TOF platform, yielding a diverse set of fragments proving to be useful for *de novo* sequencing of peptides. [41–44] Similarly, 193 nm UVPD has been used for dissociation of peptides in a wide array of bottom-up workflows.[45–49] Recently, 193 nm UVPD was applied to analysis of intact proteins, showing that the extensive fragmentation affords complete characterization of proteins.[50] This new application of 193 nm has already demonstrated to be an extremely useful tool for protein analysis[51–56], however it is still early in its use so it is unlikely that its full potential has been reached.

While delivering slightly less energy per photon (4.7 eV), 266 nm UVPD is similar to 157 nm and 193 nm in that *a*, *b*, *c*, *x*, *y*, and *z* are observed from peptide fragmentation. The major difference for 266 nm UVPD is that this wavelength is not

absorbed by the peptide backbone and is instead primarily absorbed by the aromatic side-chains of tryptophan, phenylalanine, and tyrosine residues.[57,58] Derivatization (discussed more in section 1.3 of the Introduction) of peptides with UV chromophores has extended the use of 266 nm photons.[59] 266 nm UVPD is also unique in that peptides or proteins containing disulfides bonds exhibit a homolytic cleavage of the disulfide and C-S bonds upon irradiation allowing cysteine residues to be mapped.[60] In another strategy, iodination of tyrosine residues prior to 266 nm UVPD promoted radical directed dissociation which has proven useful for probing gas-phase protein structure[61–64], for pinpointing phosphorylation sites[65], and for differentiating between D- and L amino acids in peptides.[66] Despite the recent applications using 266 nm UVPD, very little attention has focused on the pathways and distribution of fragment ions. Chapter 6 will report findings on 266 nm UVPD fragmentation as a function of charge state and peptide composition.

Finally, 351 nm (from a XeF excimer laser) or 355 nm (third harmonic of a Nd:YAG laser) UVPD has experienced more selective use in the field of proteomics compared to its counterparts. Photons from 351 nm light still deliver a moderate amount of energy (3.5 eV); however, unlike 157 nm, 193 nm, or 266 nm UVPD there is no part of the peptide which absorbs light around this wavelength, thus chromophores must be incorporated. When a peptide containing an appropriate a chromophore is exposed to 351 nm irradiation, b/y-type fragment ions are observed similar to conventional CID activation. The fact that standard peptides do not absorb 350 nm photons means that UVPD can be used a highly selective activation method for strategically-tagged peptides. For example, 351 nm UVPD has been employed to identify the antigen binding regions

of antibodies by using a cysteine selective chromophore.[67] Solvent accessibility of proteins has been evaluated using 351 nm UVPD in conjunction with a chromogenic chemical probe of lysine side-chains.[68] Recently another method has also reported that tagging of histidine and tyrosine amino acids via a chromogenic diazotisation reaction allowed bottom-up peptide analysis to be streamlined.[69] It was also shown that if a UV chromophore was attached to the N-terminus of peptides prior to exposure to 355 nm UV irradiation, the resulting MS/MS spectra contained a clean series of solely y-type ions.[70] The elimination of the b-ions was due to the fact that these fragment ions retained the UV chromophore , a factor that led to subsequent photon absorption and annihilation of the b ions, a process illustrated in **Figure 1.2.** This method showed great promise as an alternative activation for generating spectra ideal for *de novo* sequencing, and it is the subject of Chapters 3, 4, and 5.

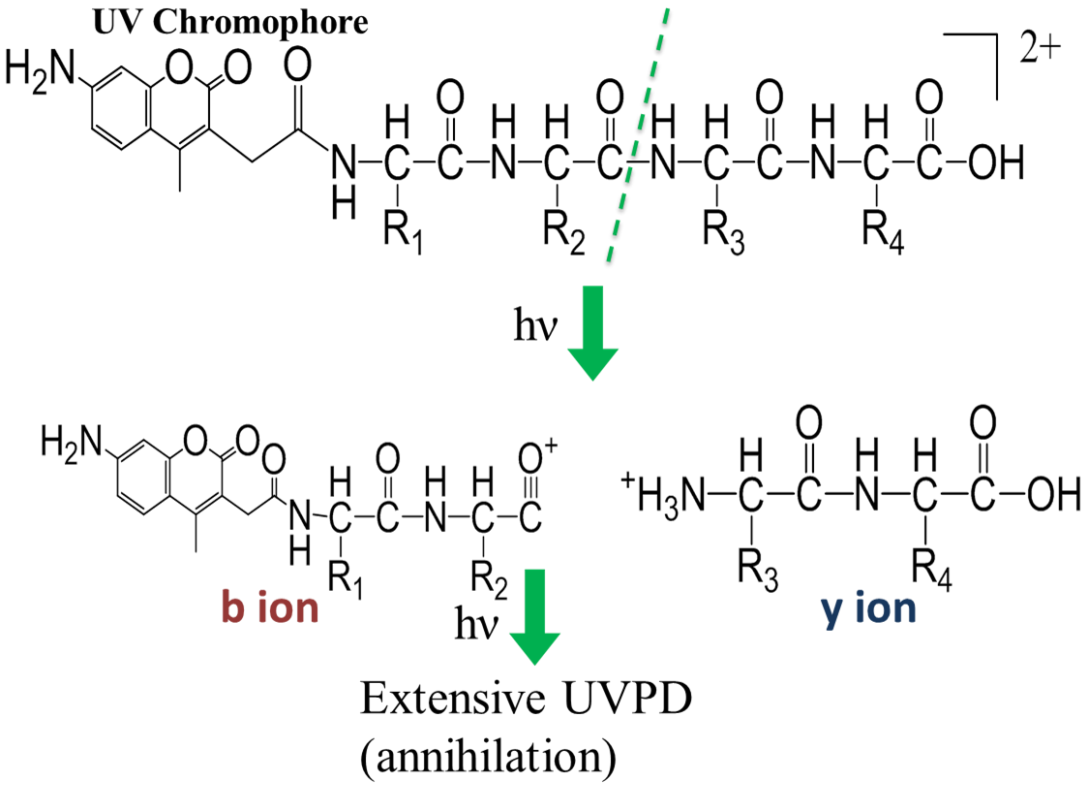


Figure 1.2: Example of 351 nm UVPD fragmentation. The b ions retain the UV chromophore making them susceptible to annihilation upon subsequent UV irradiation.

1.3 Derivatization of Proteins/Peptides

The use of chemical tags to modify functional groups via derivatization is a popular technique used to extend the versatility of many analytical methods.[71] Derivatization can be used to add charges to peptides to improve ionization or to facilitate dissociation or formation of specific types of fragment ions.[72–74] Topologies and conformations of proteins can be elucidated by tagging proteins in solution and

quantitative tracking the tags on the resulting peptides after proteolysis. [68,75–78] The 3D structure obtained from such studies is a low resolution model of proteins surface based on the reaction efficiency at particular amino acids and positions of those same amino acids in the tertiary arrangement of the protein.[79] Furthermore, incorporation of isotopic labels have proven effective for quantification. Proteins are labeled with isobaric tags that yield reporter ions upon peptide fragmentation whose abundances reflect the relative amounts of the precursor proteins.[80–82] Finally, selective tagging strategies have been used to tune fragmentation for improved *de novo* sequencing which will be addressed in section 1.3.1 of the introduction and in Chapters 3, 4, and 5.

1.3.1 Derivatization for De Novo Sequencing

For *de novo* sequencing the quality of MS/MS spectra is key to obtaining reliable results. Ideal spectra for a *de novo* type of analysis will have very few missed peptide backbone cleaves so as to give a complete series of fragment ions. The spectra would also be devoid of any type of neutral losses such as loss of water, ammonia, or amino acid side-chain cleavages as these losses are less informative or even redundant and only complicate spectra. With this goal in mind, various derivatization methods have been employed to create simplified spectra specifically for de novo sequencing. A few of these techniques target the C-terminus for derivatization. One such example is the attachment of a phosphonium reagent which biases fragmentation towards *b* ions.[83,84] In general, the vast majority of these methods target the N-terminus for modification as the primary amine presents an easy reaction site.

Derivatization methods that modify the N-terminus to create ideal spectra for *de novo* sequencing do so by modifying one series of fragment ions so that it can be easy differentiated from the other or by removing one series of ions altogether. Schemes that alter the isotopic distributions of a series of ions so they are easy to differentiate are currently the most successful example of the former. For example N-termini have been labeled with bromine tags so that when dissociation is performed any N-terminal ions will have a bromine isotope pattern making them easy to identify.[85] A different strategy has taken advantage of the more basic N-termini as the result of a Lys-N digestion as well as dimethyl isotope labeling to again allow N-terminal fragments to be easily identified due to the isotopic pattern.[86]

Despite the success of such methods the option of completely removing a series of ions from a MS/MS spectra has garnered the most attention over the years. One of the earliest reported strategies relied on modification of the N-terminus of peptides with a sulfonic acid group which caused N-terminal product ions to be neutralized when CID was performed.[87,88] Another method promoted the generation of y_{n-1} ions through amidination of the N-terminus.[89,90] Similar to the amidination of peptides, the addition of a sulfonate to the N-terminus coupled with ETD and infrared multiphoton dissociation also yield simplified MS spectra. [91–93] Acetylation has also been used as viable means to suppress b ion formation.[94] The combination of Lys-N digestion and imidazolinylation of the N-terminus of peptides resulted in formation of triplet sets of ions, when fragmented with 193 nm UVPD, creating spectra that were ideal for *de novo* sequencing.[95] Also, as discussed briefly in section 1.2.4 of the introduction, UV chromophores have been

attached to the N-terminus of peptides resulting in a single series of ions to be produced when subjected to 355 nm UVPD.[70] The numerous approaches to simplifying MS/MS spectra clearly show the advantage that can be gained from performing *de novo* sequencing in such a way, however such approaches remain underutilized in the field of proteomics. One of the major challenges facing the widespread implementation of such schemes is the lack of appropriate algorithms to exploit the unique spectra these methods produce. This dissertation will discuss the use of a *de novo* program that takes advantage of the benefits offered by such MS/MS spectra.

1.4 CHAPTER OVERVIEW:

This dissertation aims to create a *de novo* sequencing method that uses derivatization as well as the unique fragmentation generated by 351 nm UVPD. Furthermore, the use and optimization of a program that capitalizes on this unique dissociation is discussed. It will also focus on further understanding the unique fragmentation offered by different forms of UVPD.

In Chapter 3, a method is developed in which the N-terminus of peptides are modified with a UV chromophore that when exposed to 351 nm UV light causes fragmentation as well as the annihilation of *b*-ions leaving a clean series of *y*-ions that are ideal for *de novo* sequencing. The method is shown to be applicable for analysis of single proteins as well simple mixture of proteins, and finally is used to differentiate between several different green fluorescent protein (GFP) variants.

Chapter 4 improves upon this method by adapting it for analysis of complex mixtures of proteins from cell lysates. In order to take full advantage of results the design and optimization of new *de novo* sequencing algorithm was also undertaken. The effect of combining simplified fragmentation spectra from the 351 nm UVPD method and the *de novo* program resulted in a new *de novo* strategy that outperforms or is on par with current analysis software even when limited to low resolution data.

In Chapter 5 the 351 nm UVPD *de novo* method is altered to allow hybrid *de novo* searches to be performed with the use of sequencing tags and a database analysis. A UVPD/CID hybrid activation method is implemented, allowing paired spectra to be searched and enabling improved confidence of *de novo* results by taking advantages of UVPD and CID spectra together.

Chapters 4 and 5 entailed long-term collaborations with Prof. Edward Marcotte's group. In particular, Andrew Horton was responsible for the design, construction, and optimization of the de novo sequencing algorithm (UVNovo).

Finally, a fundamental study of 266 nm UVPD fragmentation was undertaken in Chapter 6. For this study all of the different charge states for six peptides were investigated in order to better understand 266 nm UVPD fragmentation based on charge state and amino acid composition. These results were compared to fragmentation from CID and 193 nm UVPD shedding light on the characteristics and pathways of 266 nm UVPD fragmentation.

1.5 REFERENCES

(1) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312* (5771), 212–217.

(2) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(4) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(5) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(6) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(7) Xu, H.; Freitas, M. A Mass Accuracy Sensitive Probability Based Scoring Algorithm for Database Searching of Tandem Mass Spectrometry Data. *BMC Bioinformatics* **2007**, *8* (1), 133.

(8) Craig, R.; Beavis, R. C. A Method for Reducing the Time Required to Match Protein Sequences with Tandem Mass Spectra. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2310–2316.

(9) Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. *Mol. Cell. Proteomics* **2007**, *6* (9), 1599–1608.

(10) De Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. Comprehensive Mass-Spectrometry-Based Proteome Quantification of Haploid versus Diploid Yeast. *Nature* **2008**, *455* (7217), 1251–1254.

(11) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(12) Standing, K. G. Peptide and Protein de Novo Sequencing by Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 595–601.

(13) Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.

(14) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(15) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(16) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(17) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(18) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(19) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(20) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.

(21) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(22) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(23) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533.

(24) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(25) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198–207.

(26) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39* (10), 1091–1112.

(27) Steen, H.; Fernandez, M.; Ghaffari, S.; Pandey, A.; Mann, M. Phosphotyrosine Mapping in Bcr/Abl Oncoprotein Using Phosphotyrosine-Specific Immonium Ion Scanning. *Mol. Cell. Proteomics* **2003**, *2* (3), 138–145.

(28) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat. Methods* **2007**, *4* (9), 709–712.

(29) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 190–208.

(30) Fernández, F. M.; Smith, L. L.; Kuppannan, K.; Yang, X.; Wysocki, V. H. Peptide Sequencing Using a Patchwork Approach and Surface-Induced Dissociation in Sector-TOF and Dual Quadrupole Mass Spectrometers. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (12), 1387–1401.

(31) Dongré, A. R.; Somogyi, A.; Wysocki, V. H. Surface-Induced Dissociation: An Effective Tool to Probe Structure, Energetics and Fragmentation Mechanisms of Protonated Peptides. *J. Mass Spectrom. JMS* **1996**, *31* (4), 339–350.

(32) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.

(33) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. Unraveling Molecular Complexity of Phosphorylated Human Cardiac Troponin I by Top down Electron Capture Dissociation/electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics MCP* **2008**, *7* (10), 1838–1849.

(34) Sweet, S. M. M.; Mardakheh, F. K.; Ryan, K. J. P.; Langton, A. J.; Heath, J. K.; Cooper, H. J. Targeted Online Liquid Chromatography Electron Capture Dissociation Mass Spectrometry for the Localization of Sites of in Vivo Phosphorylation in Human Sprouty2. *Anal. Chem.* **2008**, *80* (17), 6650–6657.

(35) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (24), 9243–9252.

(36) Crowe, M. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation (IRMPD) and Collisionally Activated Dissociation of Peptides in a Quadrupole Ion Trap with Selective IRMPD of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (11), 1581–1592.

(37) Payne, A. H.; Glish, G. L. Thermally Assisted Infrared Multiphoton Photodissociation in a Quadrupole Ion Trap. *Anal. Chem.* **2001**, *73* (15), 3542–3548.

(38) Bowers, W. D.; Delbert, S. S.; Hunter, R. L.; McIver, R. T. Fragmentation of Oligopeptide Ions Using Ultraviolet Laser Radiation and Fourier Transform Mass Spectrometry. *J. Am. Chem. Soc.* **1984**, *106* (23), 7288–7289.

(39) Hunt, D. F.; Shabanowitz, J.; Yates, J. R. Peptide Sequence Analysis by Laser Photodissociation Fourier Transform Mass Spectrometry. *J. Chem. Soc. Chem. Commun.* **1987**, No. 8, 548–550.

(40) Lebrilla, C. B.; Wang, D. T. S.; Mizoguchi, T. J.; McIver, R. T. Comparison of the Fragmentation Produced by Fast Atom Bombardment and Photodissociation of Peptides. *J. Am. Chem. Soc.* **1989**, *111* (23), 8593–8598.

(41) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 Nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19* (12), 1657–1665.

(42) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide Ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.

(43) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(44) Kim, T.-Y.; Reilly, J. P. Time-Resolved Observation of Product Ions Generated by 157 Nm Photodissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (12), 2334–2341.

(45) Moon, J. H.; Shin, Y. S.; Kim, M. S. Utility of Reaction Intermediate Monitoring with Photodissociation Multi-Stage (MSn) Time-of-Flight Mass Spectrometry for Mechanistic and Structural Studies: Phosphopeptides. *Int. J. Mass Spectrom.* **2009**, *288* (1–3), 16–21.

(46) Moon, J. H.; Yoon, S. H.; Bae, Y. J.; Kim, M. S. Dissociation Kinetics of Singly Protonated Leucine Enkephalin Investigated by Time-Resolved Photodissociation Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (7), 1151–1158.

(47) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(48) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(49) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(50) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein

Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(51) O'Brien, J. P.; Mayberry, L. K.; Murphy, P. A.; Browning, K. S.; Brodbelt, J. S. Evaluating the Conformation and Binding Interface of Cap-Binding Proteins and Complexes via Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **2013**, *12* (12), 5867–5877.

(52) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.

(53) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (37), 12920–12928.

(54) Cammarata, M.; Lin, K.-Y.; Pruet, J.; Liu, H.-W.; Brodbelt, J. Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-down Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86* (5), 2534–2542.

(55) Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D. Evolving tRNASec for Efficient Canonical Incorporation of Selenocysteine. *J. Am. Chem. Soc.* **2014**, *137* (1), 46–49.

(56) Cannon, J. R.; Martinez-Fonts, K.; Robotham, S. A.; Matouschek, A.; Brodbelt, J. S. Top-Down 193-Nm Ultraviolet Photodissociation Mass Spectrometry for

Simultaneous Determination of Polyubiquitin Chain Length and Topology. *Anal. Chem.* **2015**.

(57) Oh, J. Y.; Moon, J. H.; Kim, M. S. Sequence- and Site-Specific Photodissociation at 266 Nm of Protonated Synthetic Polypeptides Containing a Tryptophanyl Residue. *Rapid Commun. Mass Spectrom.* **2004**, *18* (22), 2706–2712.

(58) Oh, J. Y.; Moon, J. H.; Kim, M. S. Chromophore Effect in Photodissociation at 266 Nm of Protonated Peptides Generated by Matrix-Assisted Laser Desorption Ionization (MALDI). *J. Mass Spectrom.* **2005**, *40* (7), 899–907.

(59) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. Photodissociation Tandem Mass Spectrometry at 266 Nm of an Aliphatic Peptide Derivatized with Phenyl Isothiocyanate and 4-Sulfophenyl Isothiocyanate. *Rapid Commun. Mass Spectrom.* **2005**, *19* (10), 1283–1288.

(60) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(61) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358.

(62) Liu, Z.; Julian, R. R. Deciphering the Peptide Iodination Code: Influence on Subsequent Gas-Phase Radical Generation with Photodissociation ESI-MS. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (6), 965–971.

(63) Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609.

(64) Sun, Q.; Yin, S.; Loo, J. A.; Julian, R. R. Radical Directed Dissociation for Facile Identification of Iodotyrosine Residues Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2010**, *82* (9), 3826–3833.

(65) Diedrich, J. K.; Julian, R. R. Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. *Anal. Chem.* **2011**, *83* (17), 6818–6826.

(66) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(67) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(68) O'Brien, J. P.; Pruet, J. M.; Brodbelt, J. S. Chromogenic Chemical Probe for Protein Structural Characterization via Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (15), 7391–7397.

(69) Aponte, J. R.; Vasicek, L.; Swaminathan, J.; Xu, H.; Koag, M. C.; Lee, S.; Brodbelt, J. S. Streamlining Bottom-up Protein Identification Based on Selective Ultraviolet Photodissociation (UVPD) of Chromophore-Tagged Histidine- and Tyrosine-Containing Peptides. *Anal. Chem.* **2014**, *86* (13), 6237–6244.

(70) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(71) Zaikin, V.; Halket, J. M. *A Handbook of Derivatives for Mass Spectrometry*; IM Publications, 2009.

(72) Roth, K. D. W.; Huang, Z.; Sadagopan, N.; Watson, J. T. Charge Derivatization of Peptides for Analysis by Mass Spectrometry. *Mass Spectrom. Rev.* **1998**, *17* (4), 255–274.

(73) Vasicek, L.; Brodbelt, J. S. Enhanced Electron Transfer Dissociation through Fixed Charge Derivatization of Cysteines. *Anal. Chem.* **2009**, *81* (19), 7876–7884.

(74) Ko, B. J.; Brodbelt, J. S. Enhanced Electron Transfer Dissociation of Peptides Modified at C-Terminus with Fixed Charges. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (11), 1991–2000.

(75) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. Probing Protein Tertiary Structure with Amidination. *Anal. Chem.* **2005**, *77* (22), 7274–7281.

(76) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.

(77) Downard, K. M. Ions of the Interactome: The Role of MS in the Study of Protein Interactions in Proteomics and Structural Biology. *PROTEOMICS* **2006**, *6* (20), 5374–5384.

(78) Vasicek, L.; O'Brien, J. P.; Browning, K. S.; Tao, Z.; Liu, H.-W.; Brodbelt, J. S. Mapping Protein Surface Accessibility via an Electron Transfer Dissociation Selectively Cleavable Hydrazone Probe. *Mol. Cell. Proteomics MCP* **2012**, *11* (7), O111.015826.

(79) Mendoza, V. L.; Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* **2009**, *28* (5), 785–815.

(80) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–999.

(81) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K. A.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.

(82) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.

(83) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the

Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(84) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(85) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(86) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(87) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(88) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(89) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(90) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(91) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(92) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(93) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(94) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(95) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

Chapter 2

Experimental

2.1 Mass Spectrometry

All work in this dissertation entailed the use of an ion trap mass spectrometer modified for implementation of photodissociation. In many cases, peptides were derivatized prior to analysis, and these methods are described in this chapter. A Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Fisher, San Jose, CA) was used for all experiments. The dual linear ion trap configuration consists of a high pressure cell, which is used for collisional cooling of ions (allowing for more efficient trapping) and fragmentation via collision induced dissociation (CID), and a low pressure cell typically used for mass analysis. A modification to the back of the instrument allows ultraviolet photodissociation (UVPD) to be carried out in either of these two ion traps. The automatic gain control (AGC) was set to 1E4, and the heated capillary temperature was maintained at 275 °C. Standard parameters (q-value of 0.25) were used for CID. Three different types of UVPD were used in conjunction with the Velos Pro and performed in the low pressure trap. For 193 nm and 351 nm UVPD a Coherent Excistar excimer laser (Santa Clara, CA) was used with ArF (193 nm) or XeF (351 nm) as the excitation gas, respectively. For 193 nm UVPD the activation period was 4 ms (2 laser pulses), whereas the 351 nm UVPD activation period ranged from 4 ms to 50 ms (2 to 25 laser pulses). For 266 nm UVPD a Continuum Minilite Nd:YAG laser (Santa Clara, CA, USA) was used and the activation period ranged from 2 ms to 670 ms (1 to 10 laser pulses).

Samples are introduced to the mass spectrometer via electrospray ionization (ESI) which allows peptides and proteins to be gently ionized with little degradation. In order

for ionization to occur a sampled is flowed through a narrow capillary and high voltage is applied to the solution. The resulting charged spray yields multiply charged ions of the analyte of interest. For the experiments in this dissertation two types of ESI were used: traditional ESI and nanoESI. NanoESI utilizes smaller capillaries which allow for flowrates in the nL/min range and require significantly lower voltage to be applied. The lower flowrate offers better sensitivity for data collection. For traditional ESI, flowrates of 3 to 5 µl/min and a high voltage of 5 kilovolts were used to generate ions. The flowrate used for nanoESI experiments was set to 300 nL/min and the applied voltage was 2 kilovolts. Proteins and peptides were infused at concentrations between 1 and 10 µM in 49:50:1 H_2O/acetonitrile/formic acid (v/v/v) or 70/30 H_2O/acetonitrile.

2.2 LIQUID CHROMATOGRAPHY

A Dionex Ultimate 3000 nanoLC system (ThermoFisher Scientific; San Jose, CA) was used for the analysis of all complex proteolytic digests. The sample was first loaded onto a trap column via a loading pump. The trap column was packed with 5 µm Michrom Magic C18 from Bruker-Michrom (Auburn, CA) to a length of approximately 3 cm. The loading pump delivered a solution of 97.9/2/0.1 H_2O/acetonitrile/formic acid (v/v/v) at 5 µL/min for 5 min. The trap column was then placed in line with the analytical column. The analytical column was a New Objective PicoFrit capillary (Woburn, MA) (75 µm x 15 cm) packed with 3 µm Michrom Magic C18 from Bruker-Michrom (Auburn, CA). For separation, mobile phase A consisted of 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile. For analysis of single protein digests and simple mixtures 1 µg of sample was injected and the gradient for the analytical column was started at 3% B then increased to 50% B over 120 min at a flow rate of 300 nL/min. For complex mixtures (e.g. whole cell lysate) 5 µg of sample was injected and the gradient

was started at 3% B then increased to 50% B at 300 nL/min over 360 min via a slower gradient to account for increased sample complexity.

2.3 CHEMICALS

Bovine serum albumin, cytochrome C (bovine), beta-lactoglobulin (bovine), carbonic anhydrase (bovine) as well as O-methylisourea, iodoacetamide (IAM), and dithiothreitol (DTT) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Peptides ASHLGLAR and TSTEPQyQPGENL were purchased from Anaspec (Fremont, CA). ACTH (1-10, sequence SYSMEHFRWG), ACTH (1-14, sequence SYSMEHFRWGKPVG), ACTH (1-16, sequence SYSMEHFRWGKPVGKK), ACTH (1-17, sequence SYSMEHFRWGKPVGKKR), ACTH (1-24, sequence SYSMEHFRWGKPVGKKRRPVKVYP), ACTH (1-39, sequence SYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEF), were purchased from American Peptide Company (Sunnyvale, CA, USA). HPLC grade water, phosphate buffered saline (PBS), ammonium bicarbonate, and dimethyl sulfoxide were purchased from Thermo Fisher Scientific Inc. (Waltham, MA). Mass spectrometry grade trypsin-gold was purchased from Promega (Madison, WI, USA). LC-MS grade formic acid was purchased from Fisher scientific (Fair Lawn, NJ). LC-MS grade acetonitrile and water were purchased from EMD Millipore (Darmstadt, Germany). Sulfosuccinimydyl-7-amino-4-methyl-coumarin-3-acetic acid (AMCA) was purchased from Pierce Biotechnology (Rockford, IL, USA). *E.coli* lysate was graciously donated by Dr. Stephen Trent at the University of Texas at Austin.

2.4 Preparation of AMCA Modified Samples

2.4.1 Peptide Samples

For peptide reactions, 25 µL of a 1 mM peptide solution as well as 2.5 µL of 20 mM sulfo-NHS AMCA (in DMSO) were added to 270 µL of PBS. The mixture was reacted overnight in the dark at room temperature and then was cleaned up using a home packed C18 SPE cartridge. The modified peptides were eluted from the SPE cartridge with 70/30 acetonitrile/water and constituted to a final concentration of 10 µM.

2.4.1 Proteins: Guanidination

For each protein, 34 µL of a 100 µM protein solution was diluted in 150 µL of 100 mM ammonium bicarbonate prior to reduction with dithiothreitol and alkylation with iodoacetamide. Next the protein was digested with trypsin (1:20 w/w ratio) via incubation at 37 °C overnight. The digest was cleaned up using a C18 SPE cartridge and dried using a Savant DNA120 speedvac concentrator (Thermo Electron, Waltham, MA). The dried protein digests were then reconstituted in 50 µL of 7 N ammonium hydroxide prior to guanidination of the peptides. Guanidination was performed to convert the amine side-chains of lysines to unreactive homoarginine groups in a manner described previously (Reaction scheme shown in **Figure 2.1a**).[1] Briefly 15 µL of a 1 mg/mL O-methylisourea solution was allowed to react with 50 µL of 68 µM solution of the protein digest for 10 min at 65 °C. The resulting peptides were again cleaned up with a C18 SPE cartridge and dried via speedvac. The dried guanidinated digest was reconstituted in 270 µL of 1× PBS. A volume of 15 µL of 20 mM sulfo-NHS AMCA was then added and allowed to react overnight at room temperature (Reaction scheme shown in **Figure 2.1c**). The reaction was cleaned up with a C18 SPE cartridge and dried by a speedvac. Once dried, the modified digests were reconstituted in LC solvents.

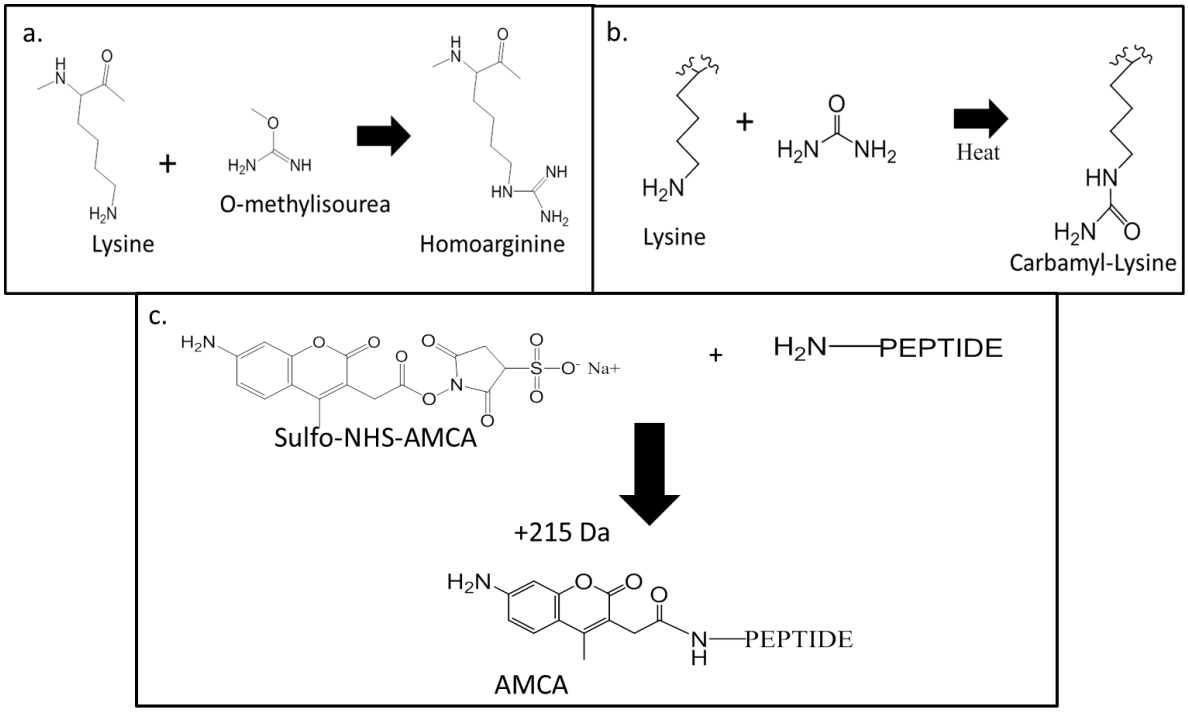


Figure 2.1: a) Reaction scheme for guanidination of lysine side chains. b) Reaction scheme carbamylation of lysine side chains. c) Reaction scheme for modification of peptide N-terminus with AMCA.

2.4.2 Proteins: Carbamylation

For model proteins and *E. coli* lysate, 50 µg of the protein or lysate in 100 µL of 50 mM sodium carbonate and 8 M urea was heated at 80 °C for 4 hours in order to block all primary amines via carbamylation to prevent reaction with AMCA (Reaction scheme shown in **Figure 2.1b**). After carbamylation, proteins were buffered exchanged into PBS to remove urea, DTT, and IAM. Proteins were digested using trypsin (1:20 w/w ratio) at 37 °C overnight. After digestion, 25 µL of 20 mM AMCA in DMSO was added to the 270 µL of the digest and incubated in the dark overnight at room temperature (Reaction scheme shown in **Figure 2.1c**). Following modification with AMCA samples were

cleaned using a C18 SPE cartridge, evaporated to dryness, and reconstituted in LC solvents.

2.5 Ultraviolet Photodissociation

For all types of UVPD used in this dissertation a Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific; San Jose, CA) was outfitted with an appropriate laser to allow for 193 nm, 266 nm, or 351 nm photons to be transmitted in the linear ion trap via a CaF_2 window (for 193 nm and 266 nm) or a fused silica window (for 351 nm) as described previously.[2]

2.5.1 UVPD at 193 nm

For 193 nm UVPD a Coherent Excistar excimer laser (Santa Clara, CA) was used. The laser was pulsed twice per scan with an energy of 2 mJ and a pulse duration of 5 ns at a frequency of 500 Hz.

2.5.2 UVPD at 266 nm

For 266 nm the Velos Pro was equipped from a Continuum Minilite Nd:YAG laser (Santa Clara, CA, USA) for UVPD. The laser was pulsed 1, 2, 5, or 10 times with an energy of 6 mJ per photon at a frequency of 10 Hz.

2.5.3 UVPD at 351 nm

For 351 nm UVPD a Coherent Excistar excimer laser (Santa Clara, CA) was used. The laser was set to 3 mJ per pulse at 500 Hz and delivered 2, 5, 10, 15, 20, or 25 pulses per scan. For LC MS/MS experiments 351 nm UVPD was performed by setting the laser to 3 mJ and 15 pulses per scan. A data dependent method was used in which the first scan was a full mass scan (m/z range of 400-2000) followed by ten MS/MS events where UVPD was performed on the ten most abundant ions from the full scan. For CID/UVPD

LC MS/MS experiments the first event was again a full mass scan (m/z 500-2000) however this was followed by alternating CID/UVPD events where CID and UVPD were performed back to back on the five most abundant peaks from the full mass scan event.

2.6 AUTOMATED DATABASE SEARCHING

2.6.1 SEQUEST

SEQUEST, operated through Proteome Discoverer v. 1.3 (ThermoFisher Scientific, San Jose, CA), was used to evaluate the improvement of sequence assignment for spectra featuring *y* ions. SEQUEST settings were altered for AMCA-modified digests so that only *y* ions were considered. AMCA was added to the database as a custom modification which could be added to the N-terminus of any peptide or any lysine. Standard settings (precursor mass tolerance 1.6 Da, fragment mass tolerance 0.8 Da, and 2 missed cleavages) were used for database searches of protein digests based on CID mass spectra; however, the SEQUEST settings were altered for the UVPD data sets. The International Protein Index (IPI) bovine database was used as the database for the SEQUEST searches.

In order to obtain a list of true peptides from the *E.coli* lysate SEQUEST, run through Proteome Discoverer v. 1.4 (ThermoFisher Scientific, San Jose, CA) was used. AMCA was added as a fixed modification on the N-terminus of all peptides. SEQUEST was set up with standard settings except that lysine cleavage of trypsin was disabled by selecting Trypsin(R) in the settings. The reason lysine cleavage was disabled is because carbamylation of lysines prevents trypsin from cleaving at lysines. The search was set so that *b*-ions would not be considered and only *y*-ions would be searched for UVPD data sets. The database used was an *E. coli* reference database from UniProt. CID/UVPD

E.coli datasets were also searched using SEQUEST with all parameters the same except both *b* and *y* ions were searched to ensure proper results for CID.

2.6.2 PEAKS

PEAKS 6 (Bioinformatics Solution Inc.; Waterloo, ON, Canada) was used for de novo sequencing of the protein digests. Raw files were converted into mzXML and then imported into the PEAKS database. Standard ion trap search parameters were used for both de novo and database comparison sections. The AMCA modification was added to the PTM section of the database in two ways: as a modification at the N-terminus and as a modification at lysines. The modification at lysine was added to ensure that if a lysine was not guanidinated, the peptide would not give false identification of an AMCA modification at the N-terminus. After completion of the PEAKS processing, the results were then checked against the International Protein Index (IPI) bovine database, allowing identification of peptides corresponding to the proteins.

2.7 REFERENCES

1. Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. Anal. Chem. **2009**, 81 (9), 3645–3653.

2. Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. Anal. Chem. **2009**, 81 (19), 8109–8118.

Chapter 3

De Novo Sequencing of Peptides Using Selective 351 nm Ultraviolet Photodissociation Mass Spectrometry[1]

3.1 Overview

Although in silico database search methods remain more popular for shotgun proteomics methods, *de novo* sequencing offers the ability to identify peptides derived from proteins lacking sequenced genomes and ones with subtle splice variants or truncations. Ultraviolet photodissociation (UVPD) of peptides derivatized by selective attachment of a chromophore at the N-terminus generates a characteristic series of y ions. The UVPD spectra of the chromophore-labeled peptides are simplified and thus amenable to de novo sequencing. This method resulted in an observed sequence coverage of 79% for cytochrome C (eight peptides), 47% forβ-lactoglobulin (five peptides), 25% for carbonic anhydrase (six peptides), and 51% for bovine serum albumin (33 peptides). This strategy also allowed differentiation of proteins with high sequence homology as evidenced by *de novo* sequencing of two variants of greenfluorescent protein

3.2 Introduction

Technological advances in tandem mass spectrometry, including improvements related to sensitivity, high resolution and high mass accuracy performance metrics, quantitative abilities, and ion activation methods, have accelerated the routine

[1] Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.
(Scott A Robotham preformed mass spectrometry experiments, sample preparation, and wrote the article. Christien Kluwe created and provided green fluorescent protein samples. Joe R Cannon helped preform data analysis of mass spectrometry results. Andrew Ellington supervised the creation of the protein samples. Jennifer S Brodbelt supervised the mass spectrometry experiments.)

identification of proteins and applications to large scale proteomics.[1] The most common strategy is the bottom-up approach which entails proteolytic digestion of proteins prior to chromatographic separation of the resulting peptides and identification based on their MS/MS fragmentation patterns. The MS/MS patterns are matched to predicted tandem mass spectra from a protein database (an *in silico* method) or directly interpreted based on the masses and mass differences of the fragment ions (*de novo* sequencing). The former method has been the dominant one used because of the tremendously powerful algorithms developed to exploit the enormous amount of genomic sequence information available. The database search algorithms include SEQUEST,[2] Mascot,[3] MassMatrix,[4] OMSSA,[5] X!Tandem,[6] among others. Despite their widespread popularity, these *in silico* database search algorithms are limited to the identification of proteins from organisms with sequenced genomes and often yield less successful results for proteins incorporating post-translational modifications.[7] A similar deficiency arises for identification of splice variants, mutants, and truncated proteins.

The shortcomings of the general database search approach has spurred the development of *de novo* sequencing methods[8–10] which offer greater flexibility for mapping unexpected post-translational modifications and sequence mutations. There have been many *de novo* programs developed to facilitate interpretation of peptide MS/MS spectra and identification of proteins, including PEAKS[11], MSnovo[12], Lutefisk,[13] DACSIM,[14] NovoHMM,[15] PepNovo,[16] EigenMS,[17] and Vonode,[18] among others Since *de novo* sequencing relies on the direct interpretation of the peptide fragmentation

patterns to assign sequences, the ideal MS/MS spectra should exhibit few missed cleavages while affording consistent series of fragment ions without extensive secondary neutral losses of water, ammonia, or side-chain groups. Some of the algorithms perform optimally for more simplified MS/MS patterns that do not contain redundant ions series, such as solely N-terminal *b* ions or C-terminal *y* ions, and thus utilize only a single ion type to construct sequence tags. Others use all ion types in conjunction with high mass accuracy information to create more elaborate connections of sequence tags[18,19]

With respect to generation of diagnostic fragmentation patterns of peptides, collisional induced dissociation (CID) is the gold standard for producing spectra containing primarily *b* and *y* ion series. [20,21] Other methods such as electron-based activation methods (electron capture dissociation ECD[22] and electron transfer dissociation ETD[23,24]), surface induced dissociation (SID),[25] and photodissociation (PD)[26–29] have gained traction in recent years to complement CID and overcome some of its limitations. For example, the electron-based methods are well-known for facilitating mapping of PTMs. To compensate for incomplete ion series in CID spectra during *de novo* sequencing applications, ETD spectra have been integrated with CID spectra to improve sequencing accuracy.[30] Both SID and PD afford richer MS/MS spectra than CID, and in particular UVPD results in informative *a/x* fragmentation patterns for peptide anions, a particular shortcoming of CID.[31] Each UV photon deposits considerable energy (i.e. 157 nm is 7.9 eV, 193 nm is 6.4 eV, 266 nm is 4.7 eV, and 351 nm is 3.5 eV), and the activation process is fast (5-10 ns pulse) and efficient, as illustrated by a number of

groups.[31–49] Reilly *et al.* has explored UVPD at 157 nm to generate rich fragmentation patterns that have been used for *de novo* sequencing.[50,51] UVPD at 193 nm similarly results in extensive peptide backbone cleavages that lead to a large array of fragment ions.[31–37] We have integrated 193 nm UVPD with Lys-N proteolysis for *de novo* sequencing based on the formation of series of triplet N-terminal ions (*a,b,c*) .[52]

Since many of the *de novo* algorithms exhibit superior performance for MS/MS spectra containing well-defined fragment ion series, there have also been numerous efforts to tailor ion activation to optimize peptide dissociation. Most activation methods of native peptides give both N-terminal and C-terminal ions,[20,21,53,54] but strategic derivatization methods have been developed to enhance specific series of fragment ions or eliminate others.[48,50,52,55–72] For example Keough *et al.* were one of the first groups to show that derivatizing the N-terminus of peptides with an ionizable sulfonic acid group resulted in neutralization of N-terminal product ions formed during CID (for singly charged precursors), thus leaving only the C-terminal *y* series as detectable charged fragment ions.[55,56] The success of simplifying *de novo* sequencing by eliminating a complete ion series motivated others to adopt and further refine this approach. In this context, Reilly *et al.* developed an N-terminus amidination procedure that promoted the production of y_{n-1} ions.[60,61] Our group has adapted a similar N-terminal sulfonation method for use with alternative activation methods including ETD and IRMPD.[62–64] Lee *et al.* demonstrated that isotopically-labeled sulfonation reagents allowed facile differentiation of N-terminal ions from other ion types because they appeared as light and

heavy doublets.[58] Similar to this latter method, Kim *et al.* labeled the N-terminus of peptides with a bromine tag so that these N-terminal peptides fragments could be easily distinguished from C-terminal fragments owing to the presence of the bromine isotope signature, thus allowing peptides to be sequenced based on the N-terminal fragments.[66] In another study Samgina modified the N-terminus of peptides by 2,4,6-trimethylpyridinium or acetylation in order to prevent cyclization of the resulting *b* ions and hence improve *de novo* sequencing.[67] In an effort to bias fragmentation via creation of peptides with a basic N-terminus, Boersema et al. used Lys-N to digest proteins,[68] and the resulting peptides favored formation of N-terminal fragment ions upon MS/MS. Hennrich *et al.* developed a method that also exploited Lys-N digestion in conjunction with dimethyl isotope labeling, ultimately resulting in N-terminal fragment readily differentiated from C-terminal fragments based on the isotopic labeling pattern.[69] Yamaguchi *et al.* demonstrated derivatization of the C-terminus using a fixed charge phosphonium reagent that facilitated production of *b*-type ions.[71,72] Very recently Ji *et al.* reported an on-tip derivatization method to append fixed charges to the N-termini of peptides and then integrated CID and ETD spectra to increase the confidence of the identified sequences based on more comprehensive N-terminal fragment maps.[73] In a rather different approach that harnessed the selectivity of UVPD, we reported that the attachment of a UV chromophore to the N-terminus of a peptide resulted in almost exclusive product ion of *y* ions upon UVPD at 355 nm, as the chromophore-labelled *b* ions were selectively

annihilated upon multiple UV laser pulses.[48] Similarly, modification of the C-terminus instead of the N-terminus led to detection of *a*- and *b*-type ions upon UVPD.

The ability to tailor the fragment ion series motivated our interest to adapt the chromophore-mediated UVPD approach for *de novo* sequencing in the present study. Here we report a strategy that utilizes UVPD at 351 nm to selectively activate and dissociate tryptic peptides labeled at their N-termini with chromophores. The advantage of using radiation around 350 nm is that this wavelength range is not absorbed by native peptides, thus affording high selectivity for only those peptides modified with a suitable UV chromophore. Moreover, upon photoirradiation the chromophore-labeled fragment ions (e.g. *b* ions in the present study due to the N-terminal tagging procedure) are selectively annihilated, whereas the non-absorbing *y* ions survive as stable product ions. The resulting MS/MS spectra exhibit complete series of *y* ions, making them well-adapted for *de novo* sequencing in LCMS workflows. In sum, UVPD using 351 nm photons offers two types of selectivity: (1) only the chromophore-labelled peptides respond to the photoirradiation, a factor which alleviates MS/MS contributions from other confounding background components in the sample (i.e. precursor overlap), and (2) the UVPD spectra of the chromophore-labelled peptides are simplified and highly suited to *de novo* algorithms. As demonstrated in this study, a supercharged mutant green fluorescent protein (GFP), designed to improve reversibility of unfolding[74], is confidently identified by this 351 nm UVPD *de novo* sequencing method.

3.3 EXPERIMENTAL

3.3.1 Materials

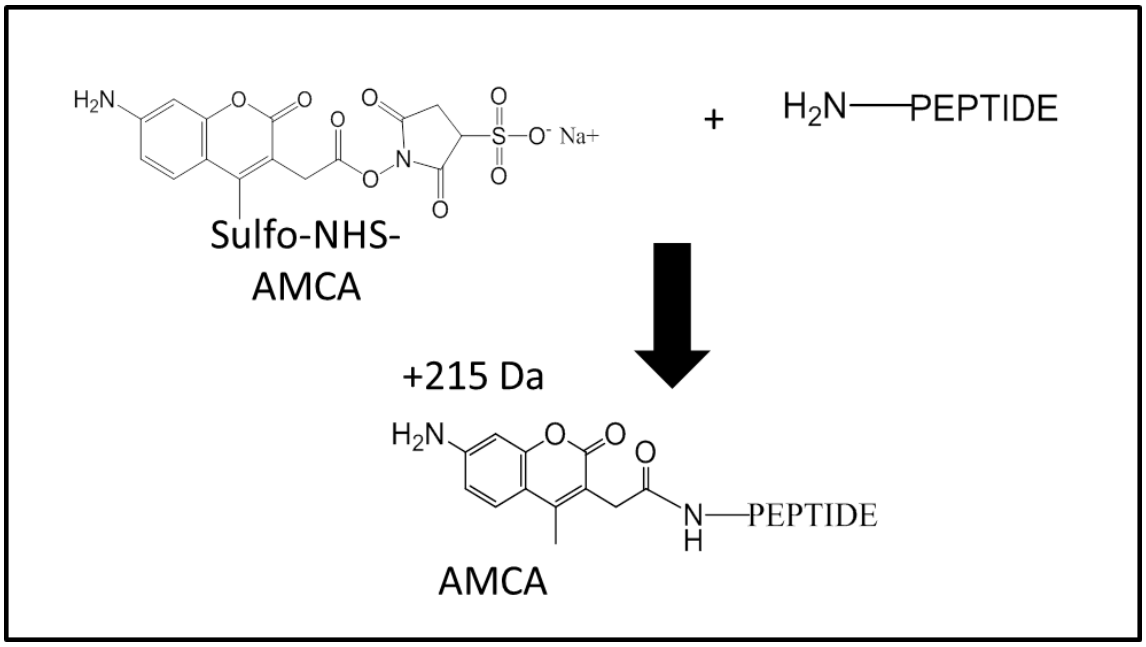
Peptides ASHLGLAR and TSTEPQyQPGENL were purchased from Anaspec (Fremont, CA, USA). The lower case "y" represents a phosphorylated tyrosine. Proteins cytochrome C (bovine), bovine serum albumin, beta-lactoglobulin (bovine), carbonic anhydrase (bovine) as well as O-methylisourea, iodoacetamide (IAM), and dithiothreitol (DTT) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile, HPLC grade water, phosphate buffered saline (PBS), ammonium bicarbonate, dimethyl sulfoxide were purchased from Thermo Fisher Scientific Inc. (Waltham, MA, USA). Sulfosuccinimydyl-7-amino-4-methyl- coumarin-3-acetic acid (sulfo-NHS AMCA or AMCA for short) was purchased from Pierce Biotechnology (Rockford, IL, USA).

3.3.2 Preparation of Mutant Green Fluorescent Proteins

Green fluorescent protein mutants were prepared in the following manner. Plasmids encoding the GFP constructs were transformed into BL21(DE3) cells, which were than plated on LB+Antibiotic. pET21 expression vector with Ampicillin resistance was used. Single colonies were inoculated into 3 mL overnight cultures. The cultures were then diluted with fresh media at a ratio of 1:200 and grown to a density of 0.6-0.8. Cells were then induced with 1 mM IPTG and incubated for 18-22 hours at 18 °C. After incubation cells were spun down and resuspended in IMAC buffer supplemented with benzonase. The cells were then lysed by sonication and lysates cleared by centrifugation at 40,000xg for 30 min. Cleared lysates were loaded onto a Ni-NTA column for affinity purification using the hexahistidine tag. Non-specific proteins were removed by washing with buffers containing increasing amounts of imidazole. GFP was then eluted with high imidazole and dialyzed 10 mM Tris, 1 mM EDTA (pH 8.0).

3.3.4 Modification of Peptides and Proteins

For peptide reactions 25 µL of a 1 mM peptide solution as well as 2.5 µL of 20 mM sulfo-NHS AMCA (in DMSO) were added to 270 µL of PBS. The mixture was allowed to react overnight in the dark at room temperature and then was cleaned up using a home packed C18 SPE cartridge. The modified peptides were eluted from the SPE cartridge with 70/30 acetonitrile/water, and constituted to a final concentration of 10 µM. For each protein, 34 µL of a 100 µM protein solution was diluted in 150 µL of 100 mM ammonium bicarbonate prior to reduction with dithiothreitol and alkylation with iodoacetamide. Next the protein was digested with trypsin via incubation at 37 °C overnight. The digest was cleaned up using a C18 SPE cartridge, and dried using a Savant DNA120 speedvac concentrator (Thermo Electron, Waltham, MA, USA). The dried protein digests were then reconstituted in 50 µL of 7 N ammonium hydroxide prior to guanidination of the peptides. Guanidination was performed to convert the amine side-chains of lysines to unreactive homoarginine groups. In order to guanidinate the digested peptides, 15 µL of a 1 mg/µL O-methylisourea solution was allowed to react with 50 µL of 68 µM solution of the protein digest for 10 min at 65 °C. The resulting peptides were again cleaned up with a C18 SPE cartridge and dried via speedvac. The dried guanidinated digest was reconstituted in 270 µL of 1x PBS. 15 µL of 20 mM sulfo-NHS AMCA was then added and allowed to react overnight at room temperature, resulting in coupling of the sulfo-NHS-AMCA to the N-terminal amine group (**Scheme 3.1**). The reaction was cleaned up with a C18 SPE cartridge and dried by a speedvac. Once dried the modified digests were reconstituted in LC solvents.



Scheme 3.1: Modification of N-terminus of peptides by AMCA.

3.3.5 MS analysis of Modified Peptides

A Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific; Waltham, MA, USA) was modified to allow introduction of a laser beam through an optical port in a manner similar to that described previously.[75] A Coherent 351 nm excimer laser (Coherent; Santa Clara, CA, USA) was used for UVPD. The laser was set to 3 mJ per pulse at 500 Hz, and the number of pulses was varied from 2 to 25 pulses to optimize UVPD. For all peptides both CID and UVPD experiments were carried out in order to compare the contributions from *b/y* ions to the MS/MS spectra. For CID a normalized collision energy (NCE) of 35% was used, with an activation time of 10 ms and a q value of 0.25. Peptides were infused by ESI at 5 ul/min in the positive mode..

3.3.6 **LCMS/MS analysis of Modified Proteins**

For modified protein digests, the same sets of UVPD and CID parameters were used as described above. HPLC separation was carried out using a Dionex NSLC 3000 nanoLC system (Thermo Scientific; Waltham, MA, USA) which was interfaced to the Velos Pro mass spectrometer. Fifteen laser pulses (at 500 Hz) were used per scan. The LC gradient consisted of mobile phase A: water with 0.1% formic acid and mobile phase B: acetonitrile with 0.1% formic acid. The loading solvent used was 98% water/2% acetonitrile with 0.1% formic acid. For each run 1 µL of the sample was injected and loaded on to a trap column prior to separation and elution from the analytical column. The column was produced in-house using a New Objective PicoFrit analytical column (Woburn, MA) (75 µm x 15 cm) and packed with 3 µm Michrom Magic C18 from Bruker-Michrom (Auburn, CA, USA). The gradient started at 3% B and increased to 50% B over 120 min at a flow rate of 300 nL/min. Companion LC-MS/MS experiments were undertaken using CID to allow comparisons to UVPD.

3.3.7 PEAKS analysis

PEAKS 6 (Bioinformatics Solution Inc.; Waterloo, ON, Canada) was used for de novo sequencing of the protein digests. Raw files were converted into mzXML and then imported into the PEAKS database. Standard ion trap search parameters were used for both de novo and database comparison sections. The AMCA modification was added to the PTM section of the database in two ways: as a modification at the N-terminus and as a modification at lysines. The modification at lysine was added to ensure that if a lysine was not guanidinated, the peptide would not give false identification of an AMCA modification at the N-terminus. After completion of the PEAKS processing, the results were then checked against the International Protein Index (IPI) bovine database, allowing identification of peptides corresponding to the proteins. In order to check the results

obtained for the mutant green fluorescent proteins, the sequences of all mutants were added to a modified IPI bovine database and subsequently searched.

3.3.8 SEQUEST analysis

SEQUEST, run through Proteome Discoverer v. 1.3 (ThermoFisher Scientific, San Jose, CA), was used to evaluate the improvement of sequence assignment for spectra featuring y ions. Standard settings were used for database searches of protein digests based on CID mass spectra; however, the SEQUEST settings were altered for the UVPD data sets acquired for the AMCA-modified digests so that only y ions were considered. The International Protein Index (IPI) bovine database was used as the database for the SEQUEST searches.

3.4 RESULTS

3.4.1 351 nm UVPD of Modified Peptides

The streamlined de novo strategy described in this report hinges on the fact that 351 nm light is not absorbed by native peptides, thus providing the basis for the selective UV photoactivation of modified peptides and preferential annihilation of the chromophore-labelled sequence ions (e.g. the N-terminal ions in the present study). Modification of the N-termini of peptides is readily accomplished via in-solution derivatization with the sulfo-NHS-AMCA reagent (**Scheme 3.1**). To illustrate the selective photodissociation of chromophore-labelled peptides, the UVPD mass spectra of unmodified and AMCA-modified peptides were compared. The comparison of one such model peptide is shown in **Figure 3.1**. From this comparison it is evident that unmodified peptides do not dissociate upon irradiation with 351 nm photons (**Figure 3.1A**), thus demonstrating the poor intrinsic absorption coefficient of peptides at this wavelength. The corresponding AMCA-modified peptide not only undergoes extensive

photodissociation (**Figure 3.1B**) but also exhibits a dominant series of y ions, but no or low abundance b ions. The location of the AMCA chromophore at the N-terminus renders the b ions highly susceptible to UVPD, thus eliminating them from the spectrum. It should also be noted that when histidines are present in the peptide sequences, low abundance internal and immonium ions are observed in some of the resulting MS/MS spectra (see H, SH, and HGL ions labeled in **Figure 3.1B**). These internal ions are only observed at lowerm/zvalues and thus do not cause a significant impediment for de novo sequencing. Such internal ions should be considered when analyzing 351 nm UVPD mass spectra, especially if lowm/zy ions are expected or they might otherwise be mis-assigned.

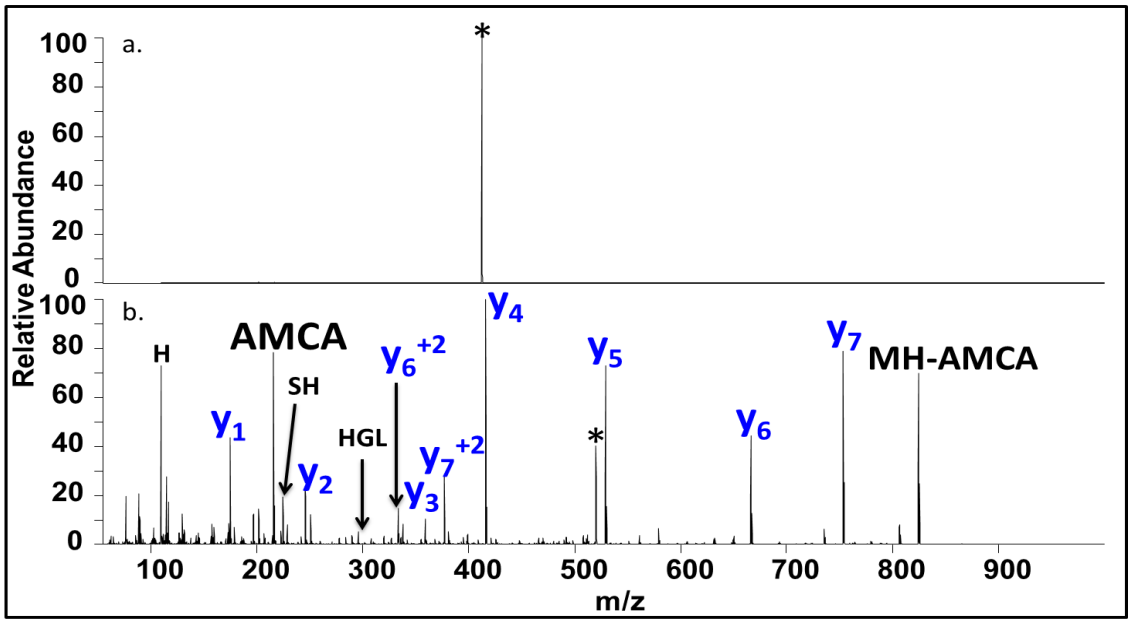


Figure 3.1: UVPD of a. ASHLGLAR (2+) and b. ASHLGLAR modified with AMCA (2+). 15 pulses at 3 mJ was used to induce fragmentation. Internal fragments H, SH, and HGL are shown. The precursor ion is labeled with an asterisk.

Implementation of the UVPD/de novo strategy on an LC time-scale required optimization of the number of laser pulses and laser power to maximize dissociation efficiency , the annihilation of b ions, and the overall throughput. To achieve this optimization, the UVPD spectra of two AMCA-modified peptides, ASGLGALR and TSETPQySPGENL, were monitored as a function of laser power and number of pulses. The outcomes of the pulse-variable experiments are shown in **Figure 3.2**. The abundances of all b and y ions from each number of pulses were tallied and used to calculate the b/y distributions as percentages in **Figure 3.2**. Even at a low number of pulses, y ions are clearly favored. As the number of laser pulses increased, the percentage of b ions decreased. For both peptides the portion of y ions plateaued after 15 pulses, and there was no significant gains using a greater number of laser pulses. In light of these results, 15 pulses were chosen for all UVPD experiments. At a laser repetition rate of 500 Hz this means that the total activation period required for 15 pulses is 30 msec. A similar procedure was undertaken to optimize the laser power, and 3 mJ was found to be ideal for producing a rich array of y ions and substantial elimination of b ions.

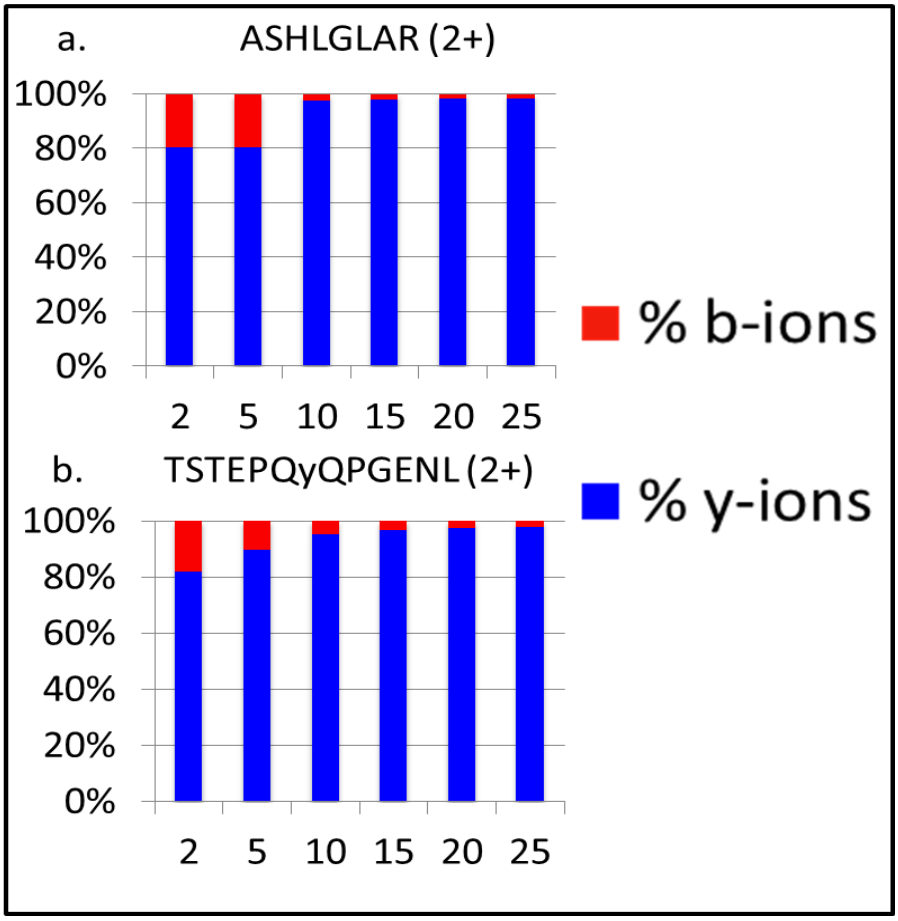


Figure 3.2: Bar graphs illustrating the decrease in the relative abundances of b ions as the number of laser pulses increase for peptides modified with AMCA. A) ASHLGLAR (2+) B) TSTEPQyQPGENL (2+)

3.4.2 De novo Sequencing of Model Proteins

The UVPD method was evaluated for de novo sequencing of tryptic peptides from selected proteins, as illustrated for bovine cytochrome c in **Figure 3.3**. In this case, cytochrome C was digested with trypsin, followed by guanidination of the epsilon-amine groups of the side-chains of lysines. Guanidination was undertaken to deactivate the lysine amines from subsequent reaction with sulfo-NHS-AMCA. The guanidination reaction efficiency was estimated to be greater than 95%. After guanidination the digest

was modified with sulfo-NHS-AMCA, then analyzed by LCMS with ion fragmentation by either CID or UVPD in order to compare the two MS/MS methods for de novo sequencing.

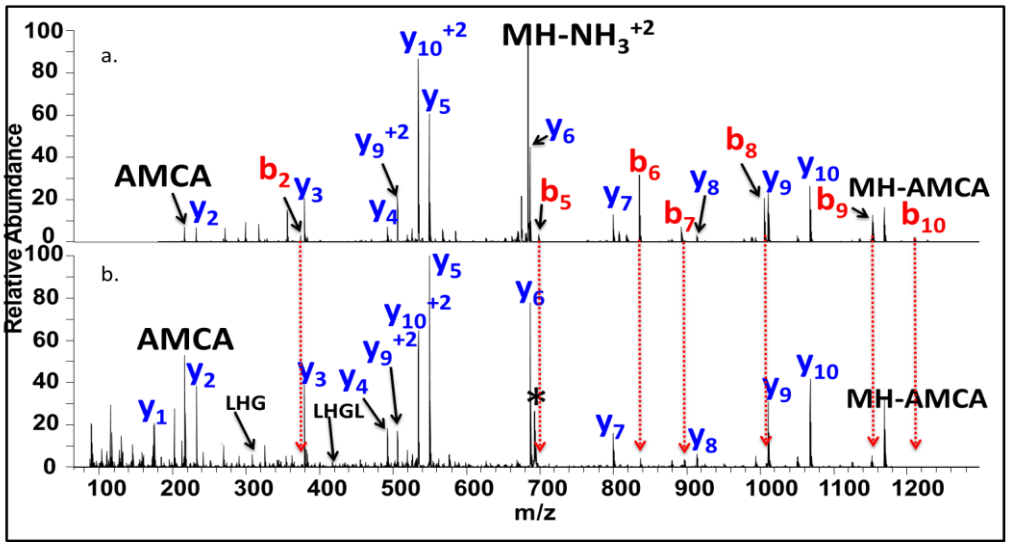


Figure 3.3: Comparison of sequence ions seen for AMCA-TGPNLHGLFGR (2+) from the cytochrome C digest by a. CID and b. UVPD (15 pulses, 3 mJ). Internal ions LHG and LHGL are shown so that the abundance of internal fragments can be seen. The precursor ion is labeled with an asterisk.

Figure 3.3 shows a comparison of the CID and UVPD spectra of one tryptic peptide TGPNLHGLFGR. The CID spectrum displays the expected series of both b and y ions (with retention of the AMCA tag on all the b ions), whereas the y ion series dominates the UVPD spectrum. Two products observed in all of the UPVD spectra of the sulfo-NHS-AMCA-modified peptides arise from elimination of the AMCA moiety (-215 Da as a neutral loss) and formation of an AMCA reporter ion (m/z 216). These two resulting ions are common to all the UVPD spectra of the AMCA-modified peptides and thus can be used to assist in pinpointing and differentiating with confidence the AMCA-modified

tryptic peptides from other species that might overlap in the chromatographic separation of complex mixtures. The addition of the AMCA modification causes one other noteworthy change: it shifts the retention time of each peptide by 10 minutes on average in comparison to the analogous unmodified peptide. This shift is caused by an increase of hydrophobicity upon addition of the non-polar aromatic chromophore to the N-terminus. This shift in retention disperses the peptides more broadly throughout the elution profile and also leads to higher ionization efficiencies in most cases due to the increasing organic content of the spray.

The AMCA modification procedure was carried out for four model proteins (BSA, carbonic anhydrase, beta-lactoglobulin, and cytochrome C) after tryptic digestion and guanidination. The resulting LC-UVPD-MS data was submitted to PEAKS6 analysis for de novo sequencing of the individual proteins. For cytochrome C eight peptides were identified through de novo sequencing, giving a sequence coverage of 79%. A total of 33 peptides were identified for BSA, accounting for a sequence coverage of 51%. Carbonic anhydrase showed a total of six peptides identified with a net sequence coverage of 25%, and five peptides were found for beta-lactoglobulin to give a sequence coverage of 47% (**Figure 3.4**). A similar LC-UVPD-PEAKS strategy was likewise

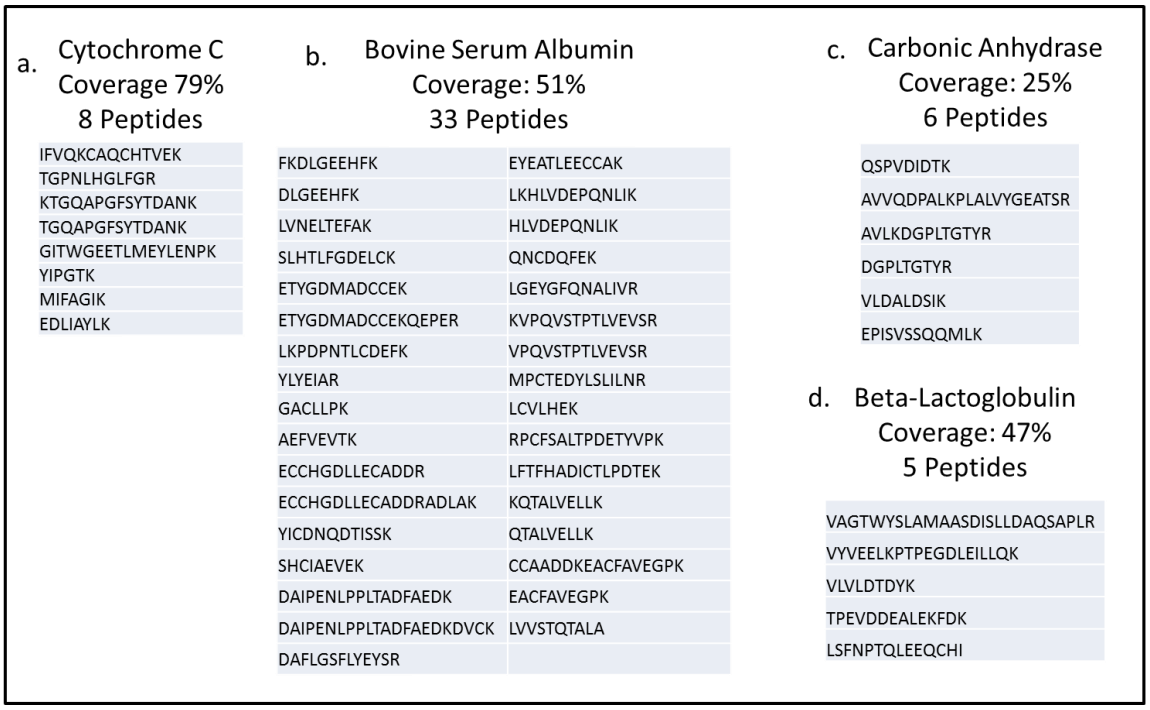


Figure 3.4: PEAKS results for individual AMCA-modified protein digests showing percent sequence coverage, number of peptides identified, as well as a list of peptides which were identified by de novo sequencing: a. cytochrome C, b. bovine serum albumin, c. carbonic anhydrase, and d. beta-lactogobulin.

undertaken for a mixture of the same four proteins. For this mixture, four peptides were found for cytochrome c, thirty for BSA, eight for carbonic anhydrase, and seven for beta-lactoglobulin which resulted in 32% sequence coverage for cytochrome C, 51% for BSA, 34% for carbonic anhydrase, and 44% for beta-lactoglobulin (**Figure 3.5**). The results from the protein mixture generally parallels those obtained for each individual protein digest; however there are a few peptides not found for both the single protein and multi-protein mixtures. The protein that shows the greatest discrepancy is cytochrome C in the protein mixture for which the number of identified peptides and net sequence coverage decrease by half compared to the cytochrome c digest alone. The number of peptides and

sequence coverage for carbonic anhydrase increases in the multi-protein mixture compared to the single protein digest. Inspection of the peptides that were not detected indicated that they were typically the lower abundance peptides that eluted close to other AMCA-labelled peptides, thus obscuring their data-dependent selection for UVPD.

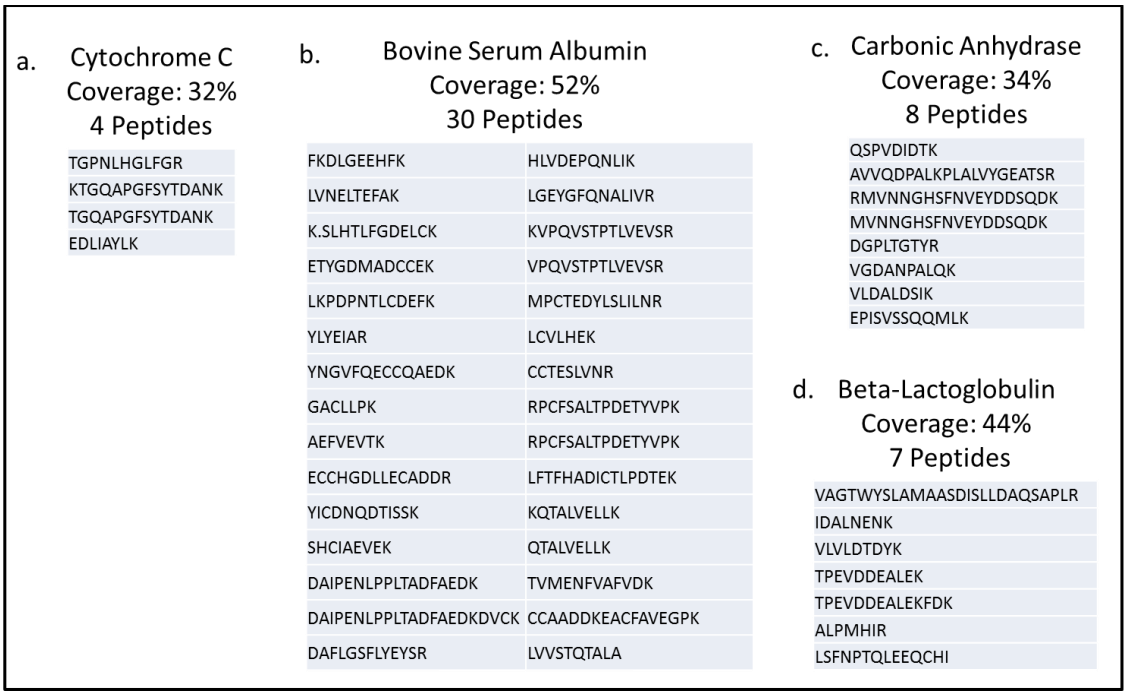


Figure 3.5: PEAKS results for multi-protein AMCA-modified protein digest. For each protein percent sequence coverage, number of peptides identified, as well as a list of peptides which were identified by de novo sequencing: a. cytochrome C; b. bovine serum albumin; c. carbonic anhydrase; and d. beta-lactogobulin.

The results from the model proteins as well as the protein mixture demonstrate the effectiveness of this method. In general, the AMCA-modified peptides all exhibited clean series of y ions upon UVPD, whereas the analogous CID spectra display both b and y ions. In two cases, better sequence coverage was obtained for the proteins when digested, modified and analyzed individually than when in a mixture. Upon further

investigation of this finding, several factors were considered and discarded as contributing to this result. At first it was suspected that the AMCA modification reaction was not tagging all the peptides in the more complex multi-protein digests, leaving some of the peptides unmodified. However, upon re-analyzing the protein mixture digests by using LC-CID-MS, it was clear that this was not the case, as very few unmodified peptides were found in the digests by CID. This observation confirmed that poor AMCA reaction efficiency for the more complex multi-protein digests was not a significant contributing factor. Further inspection of data acquired for the modified protein digests by UVPD and unmodified protein digests by CID revealed that peptides most likely to be missed in the UVPD runs were those that had low abundances in the CID runs obtained for the unmodified protein digests. This observation suggests that the multiple clean-up steps used to facilitate the AMCA-tagging strategy results in sample losses that are more detrimental for the naturally low abundance peptides.

In light of the multiple cleanup steps needed for the AMCA/UVPD strategy, the coverage obtained for AMCA-modified digests might lag that of an unmodified digest analyzed via conventional CID, thus prompting a secondary assessment of the potential analytical benefits of the former. For this assessment, the UVPD data set obtained for the AMCA-modified digest of BSA and the corresponding CID data set of unmodified BSA were evaluated by SEQUEST, a database search algorithm. SEQUEST was used because it allows MS/MS spectra to be searched selectively based solely on y ions. For this comparison, posterior error probability (PEP) scores, which indicate the probability that an observed peptide spectral match is incorrect, were determined for the MS/MS spectra generated for 33 tryptic peptides from BSA based on considering b/y ions (CID spectra) or only y ions (UVPD spectra), as summarized in **Figure 3.6**. The SEQUEST search resulted in high sequence coverage for the CID analyses as expected, with PEP scores

ranging from approximately 0.005 to 0.4. With the elimination of the b-ions from the spectra, there is a reduction in the chance that a fragment will be identified incorrectly. This results in some PEP scores that are lower (more favorable) for the UVPD data, with PEP scores ranging from 0.0001 to 0.3. In some cases the PEP scores for the UVPD spectra are several orders lower than for the corresponding CID spectra. UVPD showed improvement in the PEP scores for 73% of the identified BSA peptides. These results provide evidence that the removal of confounding b ions from the MS/MS spectra may lead to more confident assignment of peptide sequences.

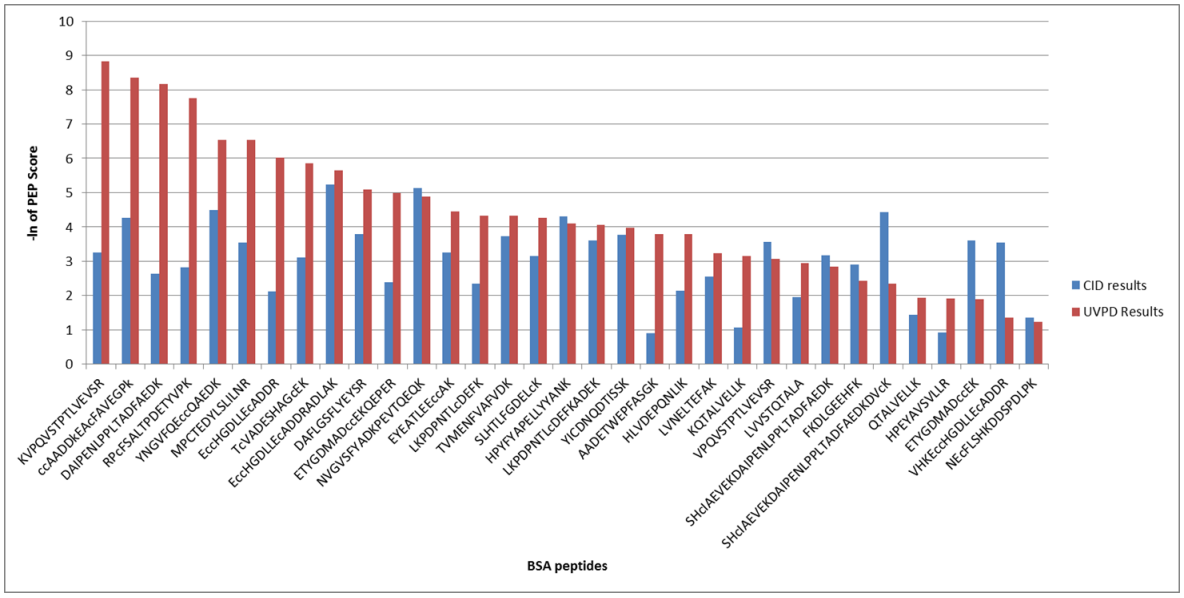


Figure 3.6: Comparison of PEP scores from SEQUEST search for trypsin digested BSA. UVPD results were scored based on only the y-ion series while CID results were scored based on b-ion and y-ion series. The negative natural log of scores is shown with higher values corresponding to higher confidant in peptide assignments. Negative natural log was used to allow for an easy comparison of results.

3.4.3 De novo Sequencing of Green Fluorescent Proteins.

To demonstrate the merits of the UVPD/de novo sequencing strategy for a biologically relevant application, the method was used to characterize wild type green fluorescent protein (GFP) and a mutant (GFP 6). GFP6 is one of fourteen different variants containing specific amino acid substitutions. Each variant has substantial sequence homology as well as unnatural sequences that make their identification difficult by conventional shot-gun strategies relying on database searches. GFP wild type (GFP WT) and mutant GFP 6 were each digested separately and modified with AMCA prior to analysis by LC-UVPD. The resulting de novo sequencing was carried out using the PEAKS program as described earlier. Seven peptides were identified from the GFP WT digest (summarized in **Figure 3.7A**), affording a sequence coverage of 29% which allowed confirmation as the wild type protein. For GFP 6, ten peptides were identified based on UVPD/de novo sequencing, returning a sequence coverage of 31% (**Figure 3.7B**). When searched against a database comprising all the GFP variants, the ten peptides led to the confident identification of the mutant as green fluorescent protein variant 6.

A sequence alignment for all of the GFPs is shown in **Figure 3.7C**. For each protein there are only a few amino acids that vary, so de novo sequencing offers the best strategy for differentiation. For the GFP WT, the threonines at positions 38 and 43 of the protein are the only two amino acids that distinguish this from the other variants except for GFP 12 (which WT differs from only by a aspartic acid at the 129 position), so pinpointing the sequence of the GEGEGDAT38NGKLT43LK and IELKGID129FK peptides are critical for identification of the protein. GFP 6 is distinguished from the other variants by only a few amino acids, including aspartic acid (D) at position 101 and the glutamic acid (E) at position 107. This necessitates satisfactory sequence coverage of

this stretch, a result achieved by 351 nm UVPD of the AMCA-modified digest and successful identification of the TISFD101DDGTYE107TR peptide.

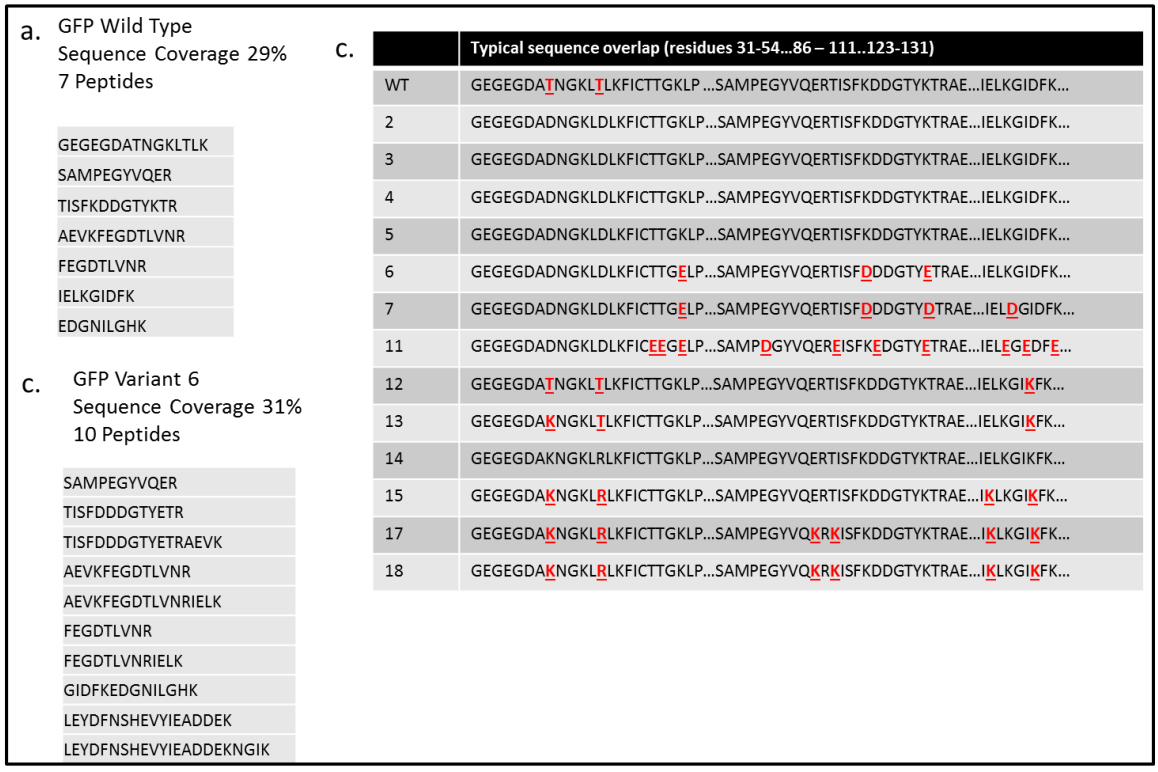


Figure 3.7: Results for AMCA-modified green fluorescent protein digests. For each protein percent sequence coverage, number of peptides identified, as well as a list of peptides which were identified by *de novo* sequencing for a. Wild Type GFP and b. GFP 6. c. Sequence alignment for all 14 GFP variants: changes to the amino acid sequence are shown in red and underlined.

3.5 CONCLUSIONS

In this study a de novo sequencing method using 351 nm UVPD and AMCA modification of the N-terminus of peptides was developed to generate a clean series of y ions to facilitate identification of proteins. Two variants of green fluorescent protein (wild type and a mutant) were differentiated from other variants based on this AMCA-tagging/UVPD-MS strategy. The multiple cleanup steps necessary for implementation of

this method led to loss of some low-abundance peptides but still allowed the identification of proteins with high sequence homology.

3.6 REFERENCES

(1) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312* (5771), 212–217.

(2) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(4) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(5) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(6) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(7) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(8) Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.

(9) Standing, K. G. Peptide and Protein de Novo Sequencing by Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 595–601.

(10) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(11) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(12) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(13) Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (11), 2594–2604.

(14) Zhang, Z. De Novo Peptide Sequencing Based on a Divide-and-Conquer Algorithm and Peptide Tandem Spectrum Simulation. *Anal. Chem.* **2004**, *76* (21), 6374–6383.

(15) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(16) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(17) Bern, M.; Goldberg, D. De Novo Analysis of Peptide Tandem Mass Spectra by Spectral Graph Partitioning. *J. Comput. Biol.* **2006**, *13* (2), 364–378.

(18) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(19) Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q. pNovo: De Novo Peptide Sequencing and Identification Using HCD Spectra. *J. Proteome Res.* **2010**, *9* (5), 2713–2724.

(20) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(21) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(22) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(23) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(24) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(25) Grill, V.; Shen, J.; Evans, C.; Cooks, R. G. Collisions of Ions with Surfaces at Chemically Relevant Energies: Instrumentation and Phenomena. *Rev. Sci. Instrum.* **2001**, *72* (8), 3149–3179.

(26) Brodbelt, J. Shedding Light on the Frontier of Photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.

(27) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(28) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(29) Brodbelt, J. S.; Wilson, J. J. Infrared Multiphoton Dissociation in Quadrupole Ion Traps. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.

(30) Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Müller, R.; Böcker, S.; Huber, C. G.; Kohlbacher, O. De Novo Peptide Sequencing by Tandem MS Using Complementary CID and Electron Transfer Dissociation. *ELECTROPHORESIS* **2009**, *30* (21), 3736–3747.

(31) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(32) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.

(33) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(34) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(35) Smith, S. I.; Brodbelt, J. S. Characterization of Oligodeoxynucleotides and Modifications by 193 Nm Photodissociation and Electron Photodetachment Dissociation. *Anal. Chem.* **2010**, *82* (17), 7218–7226.

(36) Madsen, J. A.; Cullen, T. W.; Trent, M. S.; Brodbelt, J. S. IR and UV Photodissociation as Analytical Tools for Characterizing Lipid A Structures. *Anal. Chem.* **2011**, *83* (13), 5107–5113.

(37) Hankins, J. V.; Madsen, J. A.; Giles, D. K.; Brodbelt, J. S.; Trent, M. S. Amino Acid Addition to Vibrio Cholerae LPS Establishes a Link between Surface Remodeling in Gram-Positive and Gram-Negative Bacteria. *Proc. Natl. Acad. Sci.* **2012**, *109* (22), 8722–8727.

(38) Morgan, J. W.; Russell, D. H. Comparative Studies of 193-Nm Photodissociation and TOF-TOFMS Analysis of Bradykinin Analogues: The Effects of Charge Site(s) and Fragmentation Timescales. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (5), 721–729.

(39) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of Phosphorylation Site-Specific Dissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.

(40) Yoon, S. H.; Moon, J. H.; Kim, M. S. Dissociation Mechanisms and Implication for the Presence of Multiple Conformations for Peptide Ions with Arginine at the C-Terminus: Time-Resolved Photodissociation Study. *J. Mass Spectrom. JMS* **2010**, *45* (7), 806–814.

(41) Devakumar, A.; Mechref, Y.; Kang, P.; Novotny, M. V.; Reilly, J. P. Laser-Induced Photofragmentation of Neutral and Acidic Glycans inside an Ion-Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2007**, *21* (8), 1452–1460.

(42) Zhang, L.; Reilly, J. P. Peptide Photodissociation with 157 Nm Light in a Commercial Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2009**, *81* (18), 7829–7838.

(43) Thompson, M.; Cui, W.; Reilly, J. Factors That Impact the Vacuum Ultraviolet Photofragmentation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (8), 1439–1452.

(44) Diedrich, J. K.; Julian, R. R. Site Selective Fragmentation of Peptides and Proteins at Quinone Modified Cysteine Residues Investigated by ESI-MS. *Anal. Chem.* **2010**, *82* (10), 4006–4014.

(45) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(46) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(47) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. Tyrosine Sulfation in a Gram-Negative Bacterium. *Nat. Commun.* **2012**, *3*, 1153.

(48) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(49) Parthasarathi, R.; He, Y.; Reilly, J. P.; Raghavachari, K. New Insights into the Vacuum UV Photodissociation of Peptides. *J. Am. Chem. Soc.* **2010**, *132* (5), 1606–1610.

(50) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(51) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(52) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(53) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Statistical Characterization of the Charge State and Residue

Dependence of Low-Energy CID Peptide Dissociation Patterns. *Anal. Chem.* **2005**, *77* (18), 5800–5813.

(54) Zhang, Z. Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76* (14), 3908–3922.

(55) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(56) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(57) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(58) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(59) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(60) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(61) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(62) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(63) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(64) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(65) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(66) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(67) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short

Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(68) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(69) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(70) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(71) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(72) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(73) An, M.; Zou, X.; Wang, Q.; Zhao, X.; Wu, J.; Xu, L.-M.; Shen, H.-Y.; Xiao, X.; He, D.; Ji, J. High-Confidence de Novo Peptide Sequencing Using Positive Charge Derivatization and Tandem MS Spectra Merging. *Anal. Chem.* **2013**, *85* (9), 4530–4537.

(74) Der, B. S.; Kluwe, C.; Miklos, A. E.; Jacak, R.; Lyskov, S.; Gray, J. J.; Georgiou, G.; Ellington, A. D.; Kuhlman, B. Alternative Computational Protocols for Supercharging Protein Surfaces for Reversible Unfolding and Retention of Stability. *PLoS ONE* **2013**, *8* (5), e64363.

(75) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

Chapter 4

UVnovo: A Novel *De Novo* Sequencing Algorithm Using Single Series of Fragment Ions via Chromophore Tagging and 351 nm UVPD

4.1 Overview

E.coli lysates have been modified by the use of carbamylation and the attachement of a UV chromophore to the N-terminus of digested peptides so that ultraviolet photodissociation (UVPD) can generate clean y- ion series peptides. UVPD E.coli spectra are used optimize and validate a novle *de novo* algorithm, UVnovo, which uses Random Forest model and a Hidden Markov Modle (HMM) for spectral processing allowing for ultimately offering high confidence identification of thousands of peptides from an *E. coli* lysate. UVnovo identified 47.7% of the sequences correctly when only the top result was considered and 58.2% when the top three results are considered. A filter can be applied that remove *de novo* reconstructions with low scores and thus improves results to 67.2% for the top result and 78.4% for the top three results For over half of the spectra that were identified incorrectly, the de novo results disagreed with the true sequence by two incorrect amino acids. Andrew Horton, a doctoral student in the Marcotte group is credited with the design, development, optimization, and testing of the UVnovo algorithm.

4.2 Introduction

The adoption of high throughput bottom-up mass spectrometry for proteomics has accelerated in the past decade, largely due to increasingly sophisticated peptide spectral database search methods. Although not as widely used as search engines that compare tandem MS/MS spectra to theoretical fragmentation patterns from an *in silico* digested protein database, *de novo* sequencing search algorithms have also gained popularity in

recent years.[1] *De novo* sequencing directly interprets MS/MS spectra based on both precursor mass and mass difference of fragment ions, an approach which allows identification of peptides from un-sequenced genomes. Since genomic sequence information is not required *a priori*, this method is ideal for proteomics in organisms or samples lacking genomes. In the event that gene sequences are available, *de novo* approaches are well suited for discovering unknown splice variants or other sequence polymorphisms that alter amino acid sequences, as well as for identifying peptides from unannotated parts of the genome, as in so-called proteogenomics experiments.[2] These methods are also effective at in-depth characterization of proteins *via* bottom-up analysis.

Implementation of such algorithms is far from trivial and requires "ideal" MS/MS spectra for optimal performance. Such spectra exhibit MS/MS fragmentation spanning the entire peptide sequence without missed product ions and no confounding neutral losses of water, ammonia, or side chain groups. Confidence in the *de novo* peptide match is further increased by creating spectra that contain only a single series of fragment ions[3], as multiple ion types and charge states can complicate the assignment and increase the chance of a mislabeled fragment. A variety of different ion activation methods have been implemented in order to generate these ideal MS/MS spectra. Collision-induced dissociation (CID; the gold standard of activation methods for bottom-up proteomics),[4,5] electron transfer dissociation (ETD),[6,7] and photodissociation (PD)[8–10] are all commonly used to obtain spectra for *de novo* sequencing, but all of these typically yield both N- and C-terminal fragment ions which make *de novo* interpretation more challenging.

Several programs have been designed for *de novo* sequencing of sequences of peptides based on MS/MS spectra generated by collision induced dissociation (CID) and electron transfer dissociation (ETD).[11–18] PEAKS, created by Ma *et al*.,[12] uses a four step process which involves preprocessing (noise filtering, peak centering and deconvolution), candidate computation, refined scoring, and global and positional confidence scoring. Frank and Pevzner showed with PepNovo that a scoring method which accounted for chemical and physical rules of fragmentation resulted in some of the most impressive *de novo* results to date.[14] NovoHMM, created by Fischer *et al.*, used a generative hidden Markov model (HMM) to better distinguish MS/MS signal peaks from noise.[15] Bren and Goldberg used spectral graph partitioning to create EigenMS which resulted in significant improvement in accuracy for QTOF data.[16] Mo *et al.* developed a probabilistic scoring function with a mass array-based dynamic programming algorithm to identify peptides with MSNovo.[17] Recently, Vonode, developed by Pan *et al.,* was specifically designed to take advantage of high resolution MS/MS spectra by using a unique tag scoring and customized spectral graphs for *de novo* sequencing.[18]

There has also been a great deal of effort focused on creating "ideal" spectra for *de novo* sequencing *via* chemical derivatization. Most of these methods have been implemented to either enhance or eliminate a particular fragment ion series[19–40] in the quest to obtain the most informative set of product ions for *de novo* sequencing. Keough *et al.* modified N-termini of peptides with a negatively charged sulfonic acid group prior to analysis by positive mode matrix assisted laser desorption ionization (MALDI)-MS and post source decay (PSD) fragmentation. This method predominantly produced a

series of *y* ions by silencing the N-terminal *b* ions which retained the charge neutralizing sulfonate tag.[22,23] The method was later improved by Cotter and co-workers by optimizing the 4-sulfophenyl isothiocyante (SPTIC) reaction for MALDI-PSD.[24] We have shown that the SPITC modification can also be effective in conjunction with other activation methods such as ETD and infrared multiphoton dissociation (IRMPD).[29–31] Alternative methods of fragment ion manipulation include derivatization to increase the basicity of the N-terminal primary amine, thus increasing charge localization at the N-terminus and enhancing the production of N-terminal product ions. For example, imidazolinylation of LysN metalloendopeptidase peptides resulted in the formation of N-terminal "golden sets" (a,b,c-ions) upon 193 nm ultraviolet photodissociation (UVPD).[21] Most recently we have reported a *de novo* sequencing strategy that uses 351 nm UVPD to activate peptides derivatized at their N-termini with sulfosuccinimydyl-7-amino-4-methyl-coumarin-3-acetic (AMCA) to achieve *y* ion series from tryptic peptides.[40] Furthermore, the selectivity of 351 nm UVPD ensured that only AMCA-derivatized peptides underwent photodissociation, affording spectra that were not complicated by simultaneous activation and fragmentation of co-eluting background species.

The elimination of all fragment ions except for a single ion series reduces the product ion search space, which has been shown to increase confidence for matches at both the MS and MS/MS levels.[41,42] Cannon *et al.* reported that reducing the MS1 search space by mass biased partitioning improved analyses for both middle down-sized as well as bottom up-sized peptides.[43] Our group has also shown the effective reduction of MS1 search space by using cysteine enrichment and UVPD; this strategy allowed us to

successfully analyze heavy-chain complementarity determining regions of single-chain antibodies.[44] Hansen *et al.* used amino acid composition information to reduce the MS/MS search space, thereby increasing peptide identification confidence and lowering false discovery rates.[45]

Even with the effectiveness of methods to simplify MS/MS spectra, none of these strategies have reached their full potential for *de novo* sequencing as currently there is no program able to take advantage of the benefits offered by such techniques. Here, we unite chemical derivitization and computational strategies into a new high throughput *de novo* sequencing method that combines a reduction in spectral complexity with a new program called UVnovo that uses the streamlined spectra to improve *de novo* sequencing. First, we demonstrate the near quantitative efficiency of carbamylation at capping protein lysine side-chains[46] prior to tryptic digestion. Next, selective derivatization of the N-termini of the resulting tryptic peptides *via* attachment of a UV chromophore (AMCA) prior to LC-UVPD-MS generates predominantly *y* ion MS/MS spectra. Finally, the resulting spectra are processed by UVnovo, which combines a random forest classifier and a hidden Markov model (HMM) to thousands of peptides at high confidence from an *E. coli* cell lysate.

4.3 EXPERIMENTAL

4.3.1 Materials

Mass spectrometry grade trypsin-gold was purchased from Promega (Madison, WI, USA). LC-MS grade acetonitrile and water were purchased from EMD Millipore (Darmstadt, Germany). Phosphate buffered saline (PBS), dimethyl sulfoxide were purchased from Thermo Fisher Scientific Inc. (San Jose, CA, USA)).

Sulfosuccinimydyl-7-amino-4-methyl-coumarin-3-acetic acid (AMCA) was purchased from Pierce Biotechnology (Rockford, IL, USA). *E. coli* lysate was graciously donated by Dr. M. Stephen Trent's research group at the University of Texas at Austin.

4.3.2 Modification of *E.coli* Lysate.

A schematic showing the work flow for the modification of protein samples is shown in **Figure 4.1a**. For *E. coli* lysate, 50 µg of the lysate in 100 µL of 50 mM sodium carbonate and 8 M urea was heated at 80 °C for 4 hours (**Figure 4.1b**). This caused carbamylation of all lysine side chains (ε-amines) and N-termini primary amines in order to prevent subsequent reactions with AMCA at those amines. After carbamylation, proteins were buffered exchanged into PBS to remove urea. Proteins were digested using trypsin at 37 °C overnight. After digestion, 25 µL of 20 mM AMCA in DMSO was added to the solution and kept in the dark overnight at room temperature. Following modification with AMCA samples were cleaned using a C18 SPE cartridge, evaporated to dryness, and reconstituted in LC solvents.

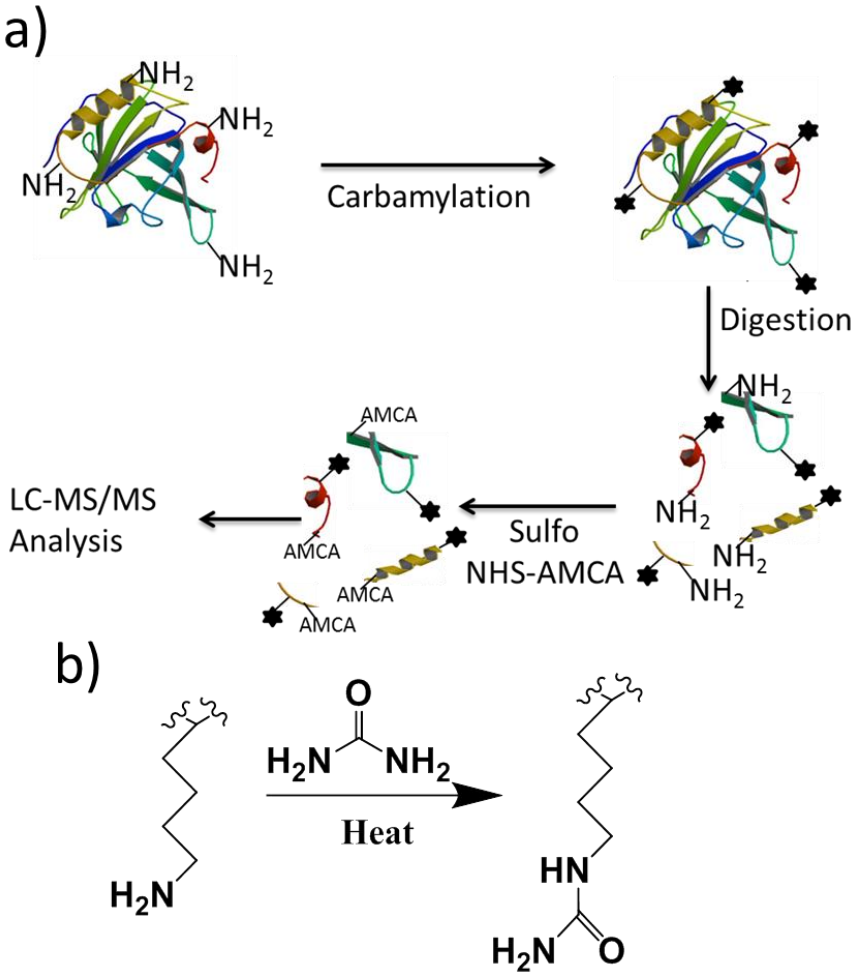


Figure 4.1: a) Workflow for carbamylation/AMCA modification, b) carbamylation reaction.

4.3.3 LC-MS/MS analysis of *E.coli* lysate

Peptides were separated and analyzed using a Dionex NSLC 3000 nanoLC system (Thermo Scientific; Waltham, MA, USA) interfaced to a Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific; San Jose, CA) modified to allow UVPD with a Coherent 351 nm excimer laser (Coherent; Santa Clara, CA, USA).[47] The laser was set to 3 mJ per pulse at 500 Hz, with 15 pulses per scan. We used a 15 cm capillary column packed with 3.5 µm particles (C18 stationary phase) with a pore size of 140 Å,

loading 5 µg of peptide mixture. Samples were eluted for analysis by UVPD using a 360 minute gradient starting at 3% B and increasing to 50% B with a flow rate of 300 nL/min; solvent A was water with 0.1 % formic acid (v/v), and solvent B was acetonitrile with 0.1% formic acid (v/v). For *E. coli* lysates the gradient was extended to 360 min to account for the increase in sample complexity.

4.3.4 SEQUEST

In order to obtain a list of high confidence peptide spectral matches, the raw spectra were analyzed using the SEQUEST and Percolator nodes of Proteome Discoverer v. 1.4 (ThermoFisher Scientific, San Jose, CA). AMCA was required as a fixed N-terminal modification, and optional oxidized methionine if any position was allowed. The precursor mass tolerance was set at ±1.6 Da due to the low resolution of ion trap spectra. SEQUEST was set up with standard settings except that lysine cleavage of trypsin was disabled by selecting Trypsin(R) in the settings. The reason lysine cleavage was disabled is because carbamylation of lysines prevents trypsin from cleaving at lysines. The search was set so that b-ions would not be considered and only y-ions would be searched for UVPD data sets. The database used was an E. coli reference database from UniProt.

4.3.5 *De novo* analysis using UVnovo

We implemented UVnovo, a *de novo* sequencing program for analysis of UVPD spectra, using the MATLAB programming language. Training and validation of UVnovo relied on spectra identified with high confidence through database searching.

4.3.5.1 Generation of high confidence data sets

Raw *E. coli* UVPD spectra were analyzed using the SEQUEST and Percolator nodes of Proteome Discoverer v. 1.4 (ThermoFisher Scientific, San Jose, CA). AMCA was required as a fixed N-terminal modification, and optional oxidized methionine of any position was allowed. The precursor mass tolerance was set at ±1.6 Da due to the low resolution of ion trap spectra. SEQUEST was set up with standard settings except that lysine cleavage of trypsin was disabled by selecting Trypsin(R) in the settings. The reason lysine cleavage was disabled is because carbamylation of lysines prevents trypsin from cleaving at lysines. The search was set so that b-ions would not be considered and only y-ions would be searched for UVPD data sets. The database used was an E. coli reference database from UniProt.

The top-ranked high confidence PSMs from charge 2+ precursor ions were used to train and benchmark UVnovo. Spectra were divided into independent training and test sets based on their respective peptide sequence, ignoring any PTM variants. 70% of identified peptides and all their associated spectra composed the training set, while the remaining 30% of peptides with associated spectra were used for validation and de novo benchmarking. Thermo *.raw files were converted to the mzXML format using MSConvert with peak picking, and peaks with an intensity < 2 were removed.

4.3.5.2 UVnovo

Following data import and preprocessing, spectral interpretation through UVnovo follows four main step:

1. Deconvolute MS^2 fragmentation peaks into predicted N-terminal peptide cleavage site locations using a random forest model.
2. Refine the cleavage site predictions using a hidden Markov model.
3. Identify amino acid sequences that best fit the predictions.
4. Score the de novo sequence reconstructions.

4.3.5.3 Deconvolution of fragmentation patters

Similar to charge deconvolution, where a predictable series of distinct peaks can be mapped to single mass, fragmentation pattern deconvolution combines the evidence from multiple related peaks into prediction of a single base peak. For example, the presence of *b*, *y*, and *y-H2O* ion peaks is strongly indicative of a specific fragmentation site along a peptide's backbone. And as with a series of charge or isotope peaks being resolved into a single 1+ charge peak, this coterie of fragment ions can be resolved into a single base peak. The process significantly cleans up a spectrum and improves interpretability.

Long the domain of descriptive, count based statistical models, the interpretation of fragmentation ion patterns can likewise be approached from a machine learning perspective. We introduce here the application of random forest models to the computational interpretation and translation of MS/MS spectra into predictions of peptide fragmentation sites.

Random forest (RF) classifiers are an ensemble machine learning method that have been successfully applied across a wide range of fields. An ensemble, in machine

learning, aggregates many individual predictive models and usually combines their output into a single prediction. RF ensembles in particular construct a 'forest' of independent decision trees. These trees are a simple type of classifier often prone to overfitting, a problem that must be carefully guarded against in most statistical models. When combined into a RF or related ensemble, however, they turn into very powerful models.

Each tree is fit on a random sample with replacement from a training data set, using a random sample of the predictors at each node in the tree. Enforcing randomness in the construction of the individual trees ensures that the trees are uncorrelated, and this reduces the problem of overfitting for the final random forest ensemble. To apply the RF to a set of input variables, each the trees in the ensemble generates a binary prediction. The random forest returns the average of these predictions as the probability of belonging to either of the two output classes.

We used MATLAB's implementation of random forests to created an ensemble of 400 decision trees. Random forests take as input a vector of predictor variables and, during training, a binary class labels for each vector. The RF ensemble was trained primarily with peak values from MS^2 spectra normalized in the following manner. Fragment m/z values were divided by a mass defect normalization constant, 1.000468, and rounded to the nearest integer. The constant was designed to minimize the residual fraction left over from rounding, following the same principle as used in Kendrick mass calculations. In the case of coincident peaks while rounding, the highest intensity peak was retained. Peak intensity values were replaced for the RF predictions by a normalized

rank score. For each spectrum, intensities were ranked from highest to lowest, and the rank was divided by the precursor mass. Local fluctuation in peak intensity was reduced by subtracting the minimum normalized rank within a sliding window.

For a fragment mass M, with M defined as the sum of amino acid masses composing the N-terminal sequence, we created a predictor vector primarily consisting of MS/MS peak scores potentially related to the mass M. This included normalized rank scores for all peaks within -50 to +63 m/z of M or its complementary C-terminal position at M_{parent}-M. Charge 2+ peaks were also included, adjusting the window locations and sizes accordingly. Additionally, 3 derived predictors were included in the predictor vector: M/M_{parent}, ceiling(M/100), ceiling((M_{parent}-M)/100). The final two are both rounded up (ceiling) and indicate which 100 Da bin the fragment ion falls in from the N and C-termini. This gave a total of 347 predictor variables at each fragmentation site.

From the training spectra described above, we created a set of positive and negative training examples for the RF modeling. The positive examples included a predictor vector for each true fragment mass M from all of the spectra. The negative examples were created similarly with a random shift applied to each of the original masses M.

Random forests can return importance measures of input variables as a side effect of their training, and we used this to perform variable selection. RF models were constructed over four rounds, and we retained a smaller set of the most important predictors for each successive iteration. There were 30 variables remaining after three rounds of selection, and these were used to train the final RF model.

This model was then be applied to the unknown spectra of the test set at each integer mass position M where a bare fragment ion was possible. This generated a new 'spectrum' of peak mass versus probability for each original spectrum. Such probabilities generated by the random forest must be treated cautiously, since RF probability estimates tend toward the extreme, 0 and 1.

4.3.5.4 Scoring of fragment sites

We implement a hidden Markov model (HMM) to compute probabilities for all potential fragmentation sites at each position within a peptide. These fragmentation sites are the hidden states of the HMM, and they represent all possible amino acid sequences with mass equal to the precursor. Therefore, the path between two states in the HMM is equivalent to an amino acid mass. Starting from 0 Da, the HMM propagates belief along all possible paths using the forward-backward procedure, a relatively efficient dynamic programming algorithm. It only keeps track of the combined probability, or posterior, for each state it reaches, regardless of how many paths coincide at the state.

This probability is primarily determined by the peak predictions from the RF, or observations. The state transition probabilities, moving from on fragment site to the next, depend in part on a first-order Markovian approximation of amino acid sequences. Our model includes the probabilities, in both the forward and backward directions, of observing a specific residue given the one that immediately preceded it. Additionally, the probability of transitioning between specific states is affected by the frequency prior of observing a specific amino acid at that position in the sequence. We have incorporated

into the HMM the expected frequencies of observing an amino acid at either of the N or C termini or as an interior residue.

These probability matrices were constructed based on amino acid frequencies in an in silico digested E. coli reference proteome. However, they could in practice be based on any sequence database similar to the organism or sample under study. The probability of observing methionine was divided between its oxidized and unoxidized forms, at a rate matching that observed in a proteomically identified E. coli peptide data set. We plan to implement an expectation maximization method for the HMM to learn automatically the most likely rate of methionine oxidation or of any other PTM not yet included. Frequency and transition probabilities were marginalized against integer residue mass, meaning the residue pairs I/L, and F/Moxidation were indistinguishable. Lysine and glutamine could be distinguished thanks to the complete carbamylation of Lys. The probability of cysteines was set to 0 since peptides were not alkylated and disulfide bonding would make observation of cysteine-containing peptides exceedingly rare.

The HMMs as described are each constructed to model a specific sequence length. As the length of a true peptide sequence is not known, HMM are constructed for the set of most likely candidate lengths of a particular spectrum. Generally, 1 to 5 models are created for each spectrum. The fragment site probabilities are different in each model since the set of sequences composing each are necessarily disjoint.

4.3.5.5 Finding the best path

The forward-backward algorithm and HMM construction described above can identify the most likely fragment mass states at each position along the sequence. These

states only represent which masses likely correspond to a fragmentation site, and they may not all derive from the same peptide. There is often not a viable amino acid transition between all of the most probable states.

We use the Viterbi algorithm on the HMM results to obtain the most likely amino acid sequence reconstruction. This moves through the fragmentation site HMM posterior probabilities and identifies the single most likely path through the set of nodes.

Our algorithmic approach is currently limited in that we only find, for a given pectrum, the single best path from each selected AA sequence length. Extending this to the proveably correct set of top *n* paths becomes a more challenging problem.

4.3.5.6 Scoring de novo reconstructions

Sequence reconstructions were scored, ranked, and filtered to create a finished set of de novo sequences. The probability that a *de novo* reconstruction is correct is equivalent to the joint probability that all predicted nodes are correct and all true nodes were predicted. As the nodes are not independent, the true probability cannot be easily derived, and a simple upper bound is chosen according to the Fréchet inequality as the probability for the single lowest node in the sequence:

$$P(\text{sequence}) \leq P(n_1 \,\&\, n_2 \,\&\, \ldots \,\&\, n_k) \leq \min(P(n_1), P(n_2), \ldots, P(n_k))$$

This score on its own can be used to rank candidate sequences. However, it is often biased in favor of the shorter of two similar sequences. Using the training set, we identified a better metric comprising a linear combination of the minimum HMM node

probability, the average node probability, and the log of the sequence length. Candidate sequences for each spectrum are ranked by this score metric. Additionally, a filter was established based on both the minimum and average node scores. Top ranked sequences that do not meet the threshold are removed along with all other sequences from the same spectra.

4.4 RESULTS

We based our strategy to increase the confidence of peptide identification on the ability to use UVPD for efficiently generating C-terminal fragment ion (*y* ion) series while eliminating N-terminal ions (*a, b* ions). This strategy required an efficient means to attach a UV chromophore at peptides' N-termini in order to target them by UVPD, while avoiding labeling of lysine side-chains that would result in indiscriminant chromophore attachment. We describe an efficient sample processing scheme that accomplishes these goals and enables UVPD-based *de novo* peptide sequencing.

4.4.1 Lysine capping with carbamylation

In order to confine AMCA modification to the N-termini of peptides, the epsilon amino group on lysine side chains must first be blocked. Previously we used guanidination, which converts lysines into homoargines via reaction with *O*-methylisourea in the presence of 7 N ammonium hydroxide.[31,40] Trypsin digestion followed by reaction in the highly basic conditions required for guanidination necessitated several clean-up steps which resulted in loss of lower abundance peptides, thus tempering the success of the method for high throughput analysis of complex

digests. Lysine capping at the intact protein level was used to circumvent this shortcoming; trypsin digest would then proceed only at arginine residues. Carbamylation was evaluated as a quick and efficient alternative to guanidination for lysine capping.[46] Heating samples at 80 °C for four hours in an 8 M urea solution resulted in complete reaction of reactive primary amines on model proteins, including the N-termini and lysine side-chains.[46] As a proof of concept, myoglobin was subjected to the carbamylation reaction, then the intact protein was directly analyzed (without proteolysis). Myoglobin has nineteen lysine residues plus a free N-terminus (**Figure 4.2a**). The resulting ESI mass spectrum showed a mass shift of 860.09 Da from the control (non-carbamylated) protein, which equates to the mass shift expected for the carbamylation of twenty amine reactive sites (see **Figure 4.2b and c**). The reaction efficiency was estimated to be nearly quantitative based on the ESI mass spectrum shown in **Figure 4.2c**.

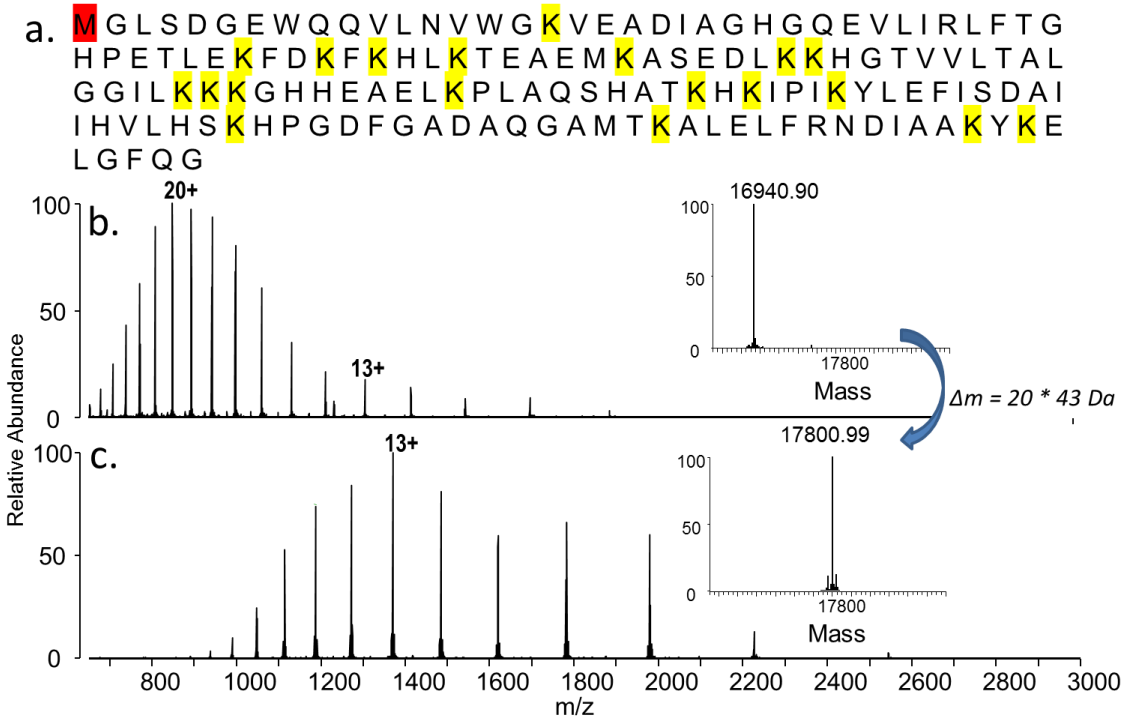


Figure 4.2: a) Myoglobin has 20 possible carbamylation sites (highlighted in yellow). Comparison of (b) the ESI-mass spectrum of myoglobin with (c) the ESI-mass spectrum of myoglobin after carbamylation in urea shows that the carbamylated myoglobin has a mass of 17,800.99 Da, corresponding to myoglobin with 20 sites of carbamylation and nearly 100% carbamylation efficiency at each Lys site. High accuracy ESI mass spectra were collected on an Orbitrap Elite.

4.4.2 Validation of UVnovo using E.coli lysate

In order to measure performance on a complex protein sample, we applied the capping/AMCA strategy to a full *E. coli* lysate. The lysate was carbamylated, digested, and derivatized with AMCA, and analyzed via LC-MS/MS with UVPD. **Figure 4.3** displays a representative UVPD spectrum from an *E.coli* peptide from elongation factor

G OS protein. As anticipated, the spectrum displays a remarkably clean series of *y* ions. The b ions (which contain the N terminus) retain the AMCA chromophore and are

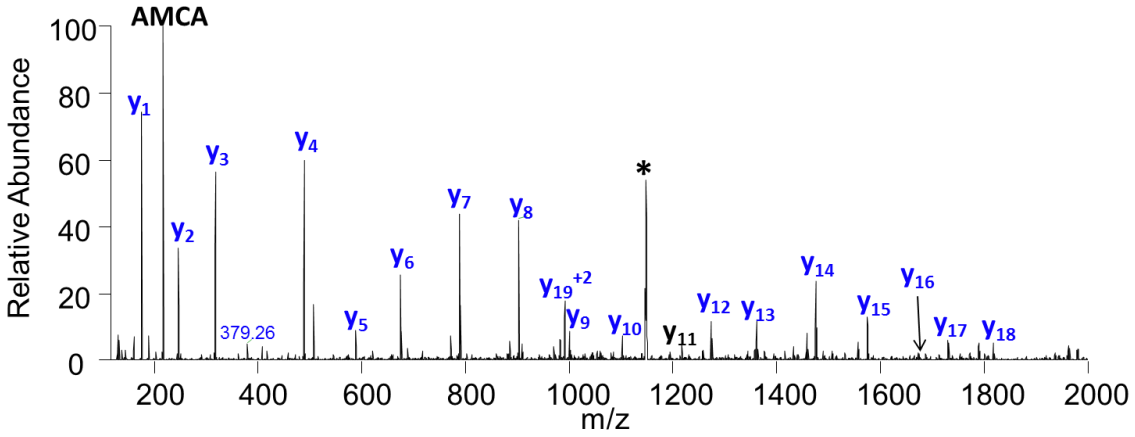


Figure 4.3: UVPD (3 mJ per pulse, 15 pulses) mass spectrum of AMCA-tagged and lysine-carbamylated VYSGVVNSGDTVLNSVKAAR (2+) from a trypsin-digested *E. coli* lysate. The precursor is labeled with an asterisk.

susceptible to photoabsorption during consecutive laser pulses. Upon increasing the number of laser pulses used for ion activation, the *b* ions are annihilated from the UVPD spectrum, leaving only the *y* series as surviving ions. Very few internal fragment ions are observed. Fragment ions are often absent C-terminal to proline, but otherwise *y* ion fragments created from backbone cleavage at all other positions are observed. The lack of complementary *b* and *y* ion pairs hampers the assignment of correct precursor mass. In all results below, the precursor mass was assigned to the integer nearest the PSM mass. However, the benefits of the UVPD method for *de novo* sequencing are twofold, and they cannot be overstated. First, with a complete *y* ion ladder, full length, gapless *de novo* reconstructions are commonly attainable for non-trivial peptides. Second, the spectra

display an ion ladder from only the C-terminus, dramatically reducing or eliminating the computationally intractable antisymmetric path problem (where mirror-image sequences propagate along both N-terminal and C-terminal ion ladders). *De novo* algorithms commonly address this problem by making imprecise assumptions, such as requiring that *b* and *y* ions not share the same integer mass. Such assumptions are unnecessary in this case.

Spectra from triplicate *E. coli* runs were processed with Proteome Discoverer SEQUEST using the Percolator node and allowing a ±1.6 Da precursor mass tolerance. Limiting the results to doubly charged precursors and top-ranked matches, 7911 high confidence identifications matching 2983 unique peptide were obtained from a total of 106,870 spectra collected across all three samples. These PSMs were divided into two independent sets, where 70% of identified peptides and all their associated spectra composed the training set. We used the remaining 30%, 2438 spectra, to test the UVnovo algorithm, thus avoiding spectrum or peptide-level biases introduced during model creation.

As illustrated in **Figure 4.4**, the UVnovo spectral processing pipeline progresses through four main steps for each MS/MS spectrum, each detailed in the Materials and Methods**.** Briefly, the spectrum is first simplified using a Random Forest (RF) classifier. At each integer mass position along the spectrum, the RF merges evidence from 30 predictors, selected and learned through training, to classify the site as either a C-terminal fragment ion or not. Next, a Hidden Markov Model is constructed from the RF spectrum and a first-order amino acid transition frequency table, and is used to estimate the

probability for each site to correspond to a true fragment ion. Using the Viterbi algorithm, the highest scoring sequence path through the network is identified. Each spectrum can then be assigned one or more potential sequences, with a single best sequence generated for each likely spectrum peptide length.

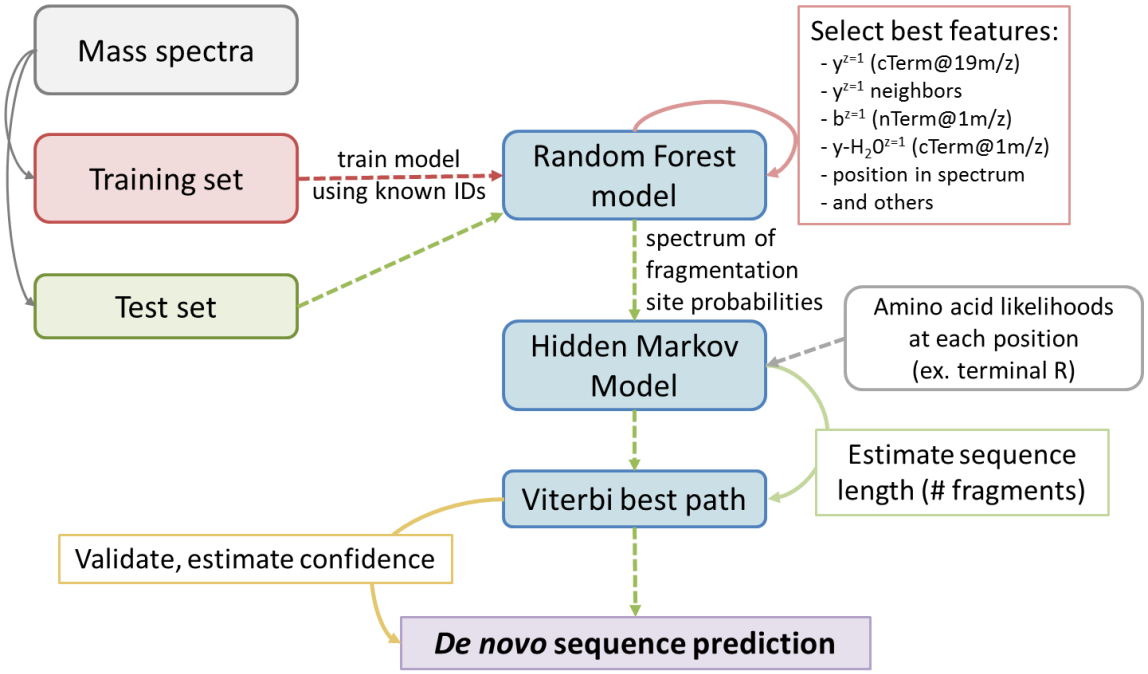


Figure 4.4: UVPD (3 mJ per pulse, 15 pulses) mass spectrum of AMCA-tagged and lysine-carbamylated VYSGVVNSGDTVLNSVKAAR (2+) from a trypsin-digested *E. coli* lysate. The precursor is labeled with an asterisk.

Having arrived at one or more de novo sequences per spectrum, the fragment node probabilities, as determined from the HMM, are used to score the sequences. The probability that a *de novo* reconstruction is correct is equivalent to the joint probability that all predicted nodes are correct and all true nodes were predicted. As the nodes are not independent, the true probability cannot be easily derived, and a simple upper bound is

chosen according to the Fréchet inequality as the probability for the single lowest node in the sequence:

$$P(\text{sequence}) \leq P(n_1 \,\&\, n_2 \,\&\, \ldots \,\&\, n_k) \leq \min(P(n_1), P(n_2), \ldots, P(n_k))$$

Candidate sequences can then be sorted in descending rank order by these scores.

For the case of the *E. coli* lysate data set, UVnovo generated correct top-ranked sequences for 47.7% of the test spectra, and when considering the top three *de novo* sequences for each spectrum, 58.2% had an exact full-length match to the corresponding SEQUEST PSM (**Figure 4.5a**). The length distribution of correct reconstructions was biased toward smaller peptides, though even >45% of peptides at least 20 residues long had a completely correct *de novo* sequence within the top three predictions (**Figure 4.5a**). As expected, the average sequence assignment precision falls with falling scores (**Figure 4.5b**). The *de novo* precision could be improved dramatically by exclusion of low ranked *de novo* reconstructions. This preferentially removed many of the false positive predictions while retaining 88.6% of true predictions, boosting the precision to 67.2% and 78.4% for top one and top three *de novo* sequence sets, respectively (**Figure 4.5c and d)**.

Such predictive accuracy—notably, seen here from relatively low-resolution mass spectra—broadly matches the performance of other *de novo* sequencing algorithms on high-resolution, higher-energy collisional dissociation (HCD) datasets. For example, a recent comparison of five *de novo* algorithms on optimized datasets showed comparable levels of accuracy.[48] The results obtained from the *E. coli* digests from UVnovo appear on par with or out-perform these other methods considering only the correct identification

from the top *de novo* match. When expanded to the top five matches the UV novo results are again on par and are only slightly behind three of the programs which boast approximately 80% correct identifications when using HCD data. A direct comparison of these algorithms on the same datasets is unfortunately not currently feasible, given the differences in fragmentation applicability.

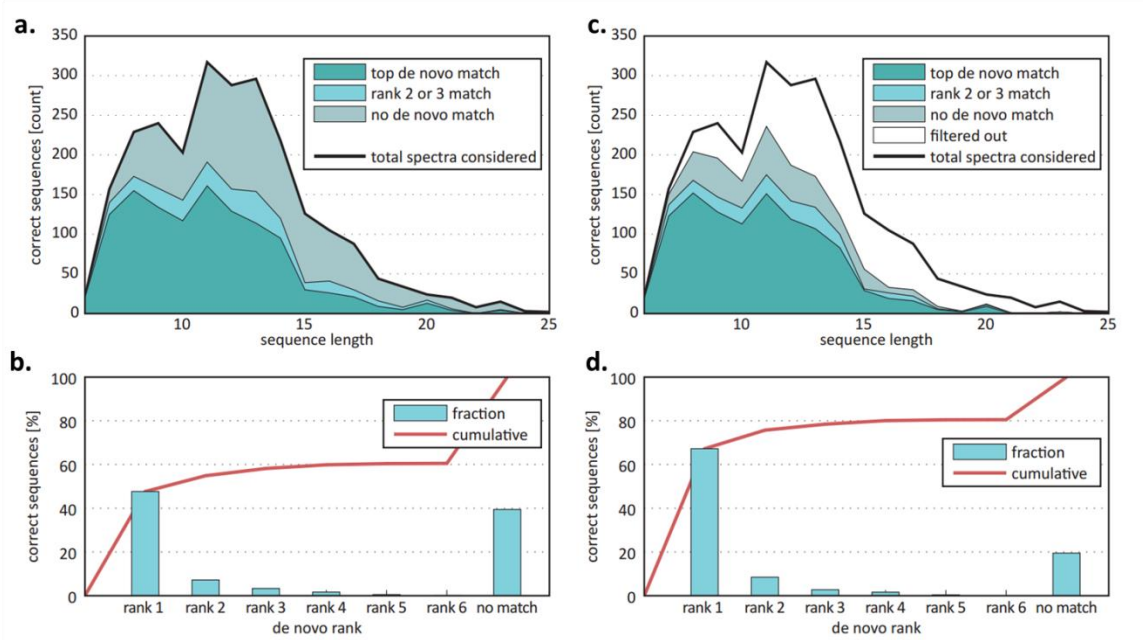


Figure 4.5: UVnovo *de novo* results for the *E. coli* lysate test set. No sequence gaps are allowed for a correct match. a) shows the number of true peptide identifications (defined as agreement with high confidence SEQUEST peptide spectral matches) by UVnovo as a function of peptide length, plotting results for the top *de novo* result and the top three *de novo* results of the same length. b) plots the percentage of total correct matches across all sequence lengths, as a function of rank, demonstrating effective ranking of hits by UVnovo. c, d) Filtering to remove low rank *de novo* reconstructions further improves relative performance.

While UVnovo boasts competitive results for full-length reconstructions, another metric is the extent of error present in incorrect sequences. **Figure 4.6** displays the

frequency and extent of amino acid sequencing errors versus peptide length. Of the misidentified sequences, half (52.0%) differ from the SEQUEST PSM by only two amino acids, suggesting that only one fragmentation site per peptide is predicted incorrectly. This, along with the high rate of full sequence recovery, demonstrates the benefit of complete fragmentation along entire peptides.

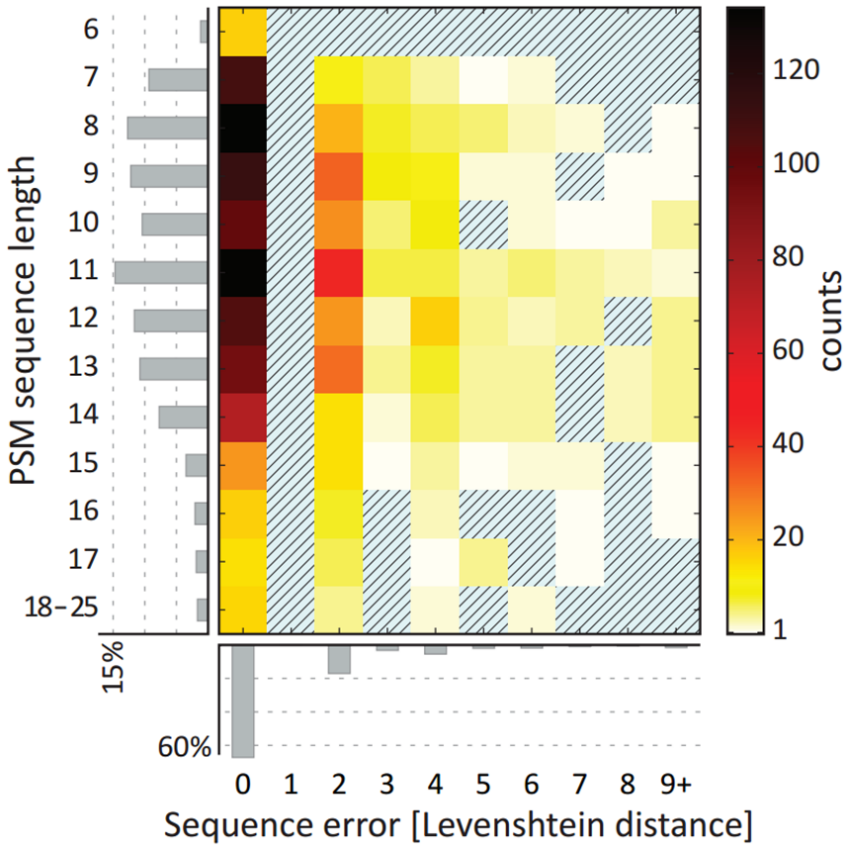


Figure 4.6: Heat map plotting relative frequency and extent of amino acid identification errors versus peptide lengths for the *E. coli* lysate test sample. Most *de novo* reconstructions are correctly reconstructed with no insertions or deletions; incorrectly identified gaps tend to be short, predominantly 2 amino acids.

We occasionally observed co-eluting peptides in our data manifesting as differences between the *de novo* sequence and SEQUEST PSM for a spectrum. In some

cases, this resulted in a hybrid *de novo* sequence blended from the two precursor peptides. Ideally, however, the differing *de novo* results complement the SEQUEST identification, and both are correct. **Figure 4.7** presents such an example, where coeluting peptides AKLLQYAER and VIELTKKAM$^{Oxy[+16]}$ both contribute to the observed spectrum and are identified through SEQUEST and *de novo*, respectively. The assignments are supported by independent identification of each from alternative, less overlapping spectra. Interestingly, the lone spectrum with a SEQUEST PSM of VIELTKKAM was partitioned into the test set, and all five AKLLQYAER spectra were part of the training set, including the spectrum in **Figure 4.7**. That this spectrum, partly composed of ions used during training, was ascribed to an unfamiliar peptide implies the UVnovo models were not significantly biased or overtrained. Additionally, this and similar observations indicate that other disagreeing *de novo* sequences may not necessarily be incorrect.

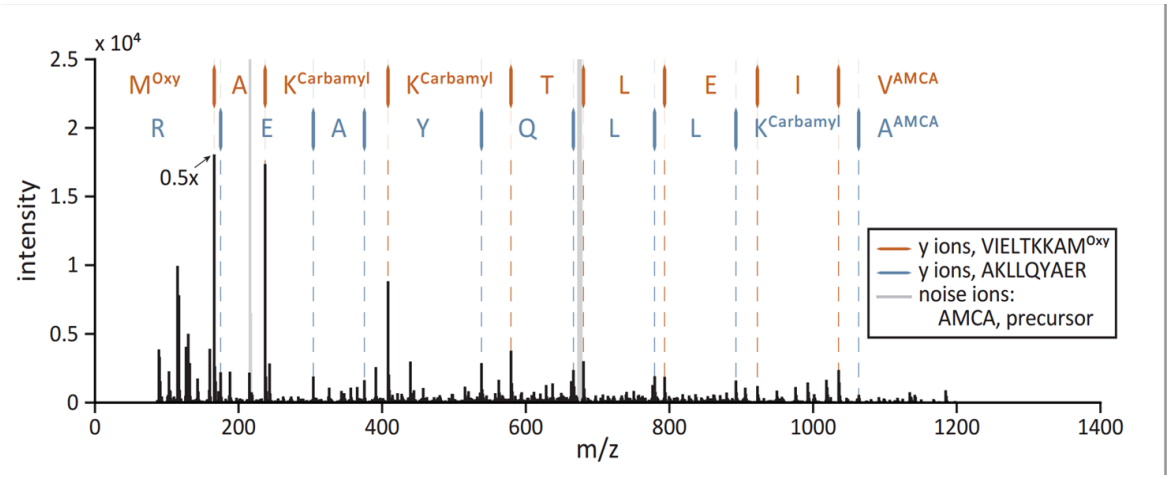


Figure 4.7: Example UVPD MS/MS spectrum in which UVnovo conflicts with SEQUEST. SEQUEST matches this spectrum to peptide VIELTKKAM$^{Oxy[+16]}$, whereas UVnovo identifies this spectrum as deriving from peptide AKLLQYAFR. The presence of a complete y-ion series for each peptide suggests the spectrum results from co-eluting peptides.

4.5 Conclusions

We describe a new experimental method and supporting algorithm, UVnovo, for *de novo* peptide sequencing by UVPD. The high efficiency of carbamylation for capping lysine side-chains at the protein level, subsequent tryptic digestion, then selective derivatization of the N-termini of the resulting peptides with a UV chromophore (AMCA) results in UVPD mass spectra that contain predominantly y ions and that are well suited for *de novo* sequencing. UVnovo combines a random forest classifier and a Hidden Markov model to simplify and interpret UVPD fragmentation spectra, enabling the *de novo* sequencing of thousands of peptides from an *E. coli* lysate at high confidence. In comparing *de novo* results to a reference set derived by database matching, UVnovo identified 47.7% of the sequences correctly when only the top result was

considered and 58.2% when the top three results are considered. These results are improved to 67.2% for the top result and 78.4% for the top three results when reasonable filters were applied to the data. When incorrect, the *de novo* results generally agreed with the true sequence within two incorrect amino acids. Further refinement of the UVnovo algorithm by capitalizing on integrated CID and UVPD methods is underway.

4.6 REFERENCES

(1) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(2) Ma, J.; Ward, C. C.; Jungreis, I.; Slavoff, S. A.; Schwaid, A. G.; Neveu, J.; Budnik, B. A.; Kellis, M.; Saghatelian, A. Discovery of Human sORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**.

(3) Noga, M. J.; Lewandowski, J. J.; Suder, P.; Silberring, J. An Enhanced Method for Peptides Sequencing by N-Terminal Derivatization and MS. *Proteomics* **2005**, *5* (17), 4367–4375.

(4) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(5) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(6) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(7) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(8) Brodbelt, J. Shedding Light on the Frontier of Photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.

(9) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(10) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(11) Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (11), 2594–2604.

(12) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(13) Zhang, Z. De Novo Peptide Sequencing Based on a Divide-and-Conquer Algorithm and Peptide Tandem Spectrum Simulation. *Anal. Chem.* **2004**, *76* (21), 6374–6383.

(14) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(15) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(16) Bern, M.; Goldberg, D. De Novo Analysis of Peptide Tandem Mass Spectra by Spectral Graph Partitioning. *J. Comput. Biol.* **2006**, *13* (2), 364–378.

(17) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(18) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(19) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(20) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(21) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(22) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(23) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(24) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(25) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(26) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(27) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(28) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(29) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(30) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(31) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(32) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(33) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(34) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(35) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide

Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(36) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(37) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(38) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(39) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(40) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(41) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. Straightforward Ladder Sequencing of Peptides Using a Lys-N Metalloendopeptidase. *Nat. Methods* **2008**, *5* (5), 405–407.

(42) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation. *Anal. Chem.* **2009**, *81* (18), 7814–7822.

(43) Cannon, J. R.; Edwards, N. J.; Fenselau, C. Mass-Biased Partitioning to Enhance Middle down Proteomics Analysis. *J. Mass Spectrom.* **2013**, *48* (3), 340–343.

(44) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(45) Hansen, T. A.; Kryuchkov, F.; Kjeldsen, F. Reduction in Database Search Space by Utilization of Amino Acid Composition Information from Electron Transfer Dissociation and Higher-Energy Collisional Dissociation Mass Spectra. *Anal. Chem.* **2012**, *84* (15), 6638–6645.

(46) Angel, P. M.; Orlando, R. Quantitative Carbamylation as a Stable Isotopic Labeling Method for Comparative Proteomics. *Rapid Commun. Mass Spectrom.* **2007**, *21* (10), 1623–1634.

(47) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of

Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

(48) Jeong, K.; Kim, S.; Pevzner, P. A. UniNovo: A Universal Tool for de Novo Peptide Sequencing. *Bioinformatics* **2013**, *29* (16), 1953–1962.

Chapter 5

Incorporation of UVPD/CID paired spectra Analysis and De *Novo* Sequence Tag /Database look-up to Expand the Capabilities of UVnovo

5.1 Overview

UVnovo capabilities are expanded to allow UVPD/CID paired spectra to be searched. *E.coli* lysate peptides modified with a UV chromophore and analyzed by ultraviolet photodissoication (UVPD) and collision induced dissociation (CID) on the same precursors to generate paired UVPD/CID spectra. Top ranked *de novo* reconstructions correctly identified 68.3% of searched spectra (compared to 62.9% for UVPD and 35.3% for CID alone) and 79.1% when the top three scored sequences from UVnovo were considered (51.9% for CID and 75.9% for UVPD alone). These results improved to 78.2% (75.9% for UVPD and 51.9% for CID alone) for the top result and 88.1% (86.7% for UVPD and 71.3% for CID alone) for the top three results when a filter was applied. The paired UVPD/CID datasets integrated with the UVnovo algorithm capitalized on the benefits of both spectra types while ignoring or canceling out the negative effects of CID and/or UVPD spectra alone (redundant ions, precursor mass assignment). UVnovo was also modified to allow performance of sequence tag de novo sequencing which can be used to identify matching sequences from a database. UVnovo identified 2,260 unique peptides compared to 1,689 identified with PEAKS and 1,957 identified with SEQUEST for an *E. coli* lysate. Andrew Horton, a doctoral student in the Marcotte group is credited with the development, optimization, testing, and modifications of the UVnovo algorithm.

5.2 INTRODUCTION

The ability of Identify proteins in complex mixtures is a challenging task in the field of proteomics. While advancements in mass spectrometry (MS) have led to great improvement in number and confidence in the proteins identified based on analysis of proteolytic peptides, the vast amount of data generated from high throughput LCMS runs makes manual searches for peptide hits impractical. As a result much effort has been directed towards the development of peptide sequencing algorithms. There are two main types of programs: *in silico* database search methods and *de novo* sequencing. *In silico* search programs match experimental MS/MS spectra of peptides to predicted tandem mass spectra generated from a protein database and have been tremendously successful in the field of cellular proteomics.[1–6] *De novo* sequencing algorithms, on the other hand, directly interpret peptide sequences from the masses and mass differences of fragment ions observed in tandem mass spectra and thus do not require the use of a protein database. The use of the latter programs have been shown to be effective for the identification of truncated proteins, post-translational modifications (PTMs), and sequence mutations, all of which pose problems for *in silico* programs.[7] The most popular de novo methods to date are PEAKS[8], MSnovo[9], PepNovo[10], pNovo[11], and Uninovo.[12]

One of the main factors that improve the success of *de novo* programs is the creation of MS/MS spectra which contain few missed peptide backbone cleaves and are devoid of any type of neutral losses such as loss of water, ammonia, or amino acid side change. Spectra that meet these conditions are ideal for *de novo* sequencing as the algorithms can derive fragment ion identities (i.e. amino acid compositions) while

limiting incorrect assignments as a result of spectrum being cluttered with redundant or uninformative ions, ones that confound *de novo* sequencing. In an effort to create these optimal mass spectra, a variety of methods have been developed based on chemical derivatization to manipulate the fragmentation patterns of peptides. Keough *et al.* made one of the first major contributions by using a negatively charged sulfonic acid group to modify the N-termini of peptides, thus resulting in the neutralization of N-terminal product ions while leaving only the C-terminal ions as detectable charged fragment ions.[13,14] This strategy proved to be an effective means to simplify the MS/MS spectra of peptides. Over the years other groups have improved upon or used similar ideas to modulate MS/MS spectra to improve *de novo* sequencing.[15–26] Some of the more recent approaches integrated Lys-N proteolysis with dimethyl isotope labeling to easily distinguish N-terminal ions from C-terminal ions[27] or derivatized the C-terminus of peptides using a phosphonium reagent to facilitate production of b-type ions.[28,29]

For the majority of bottom-up proteomics workflows, collisional induced dissociation (CID) is the dominant ion activation method, resulting in b and y fragment ions.[30,31] Other activation techniques have been developed as a means to provide greater sequence coverage and/or enhance the retention of PTM information, such as electron capture dissociation (ECD)[30], electron transfer dissociation (ETD)[31,32], and photodissociation (PD).[33–35] The use of ETD spectra for peptide sequencing has become so common that PEAKS has incorporated a *de novo* algorithm created by Liu et al. adapted for ETD spectra.[36] ETD has likewise been combined with some of the derivatization methods described above to further extend the feasibility of ETD for de

novo strategies. For example, our group reported appending a 4-sulfophenyl isothiocyante (SPTIC) modification to the N-termini of peptides to generate *de novo* friendly spectra containing mainly *z*-type ions.[37] More recently Richards *et al.* showed that ETD spectra of neutron-encoded stable isotope labeled peptides yielded $^{\bullet}$z- type doublets ions that were well-suited for *de novo* interpretation.[38] PD, specifically ultraviolet PD (UVPD), has also been shown to be an effective method to generate fragments for *de novo* sequencing.[39–44] Zhang and Reilly first reported that 157 nm UVPD produced rich fragmentation patterns of peptides that could be used for de novo sequencing.[39,40] 193 nm UVPD coupled with Lys-N proteolysis was shown by our group to yield triplet series of N-terminal ions (a,b,c) ideal for sequencing of unknown peptides.[41] We also showed that 355 nm UVPD combined with N-terminal derivatization of peptides resulted in the elimination of b-ions from tandem mass spectra resulting in a clean series of y-ions.[42] Further adaptation of this method for 351 nm UVPD proved to be very effective for de novo sequence of tryptic digest of proteins in which the peptides were tagged at the N-terminus with sulfosuccinimydyl-7-amino-4methyl-coumarin-3-acetic (AMCA).[43] Most recently, we have modified an *E.coli* lysate via carbamylation and AMCA derivatization in order to create and validate a new program to effectively perform *de novo* sequencing on 351 nm UVPD data called UVnovo.[45]

Recently de novo methods have begun to incorporate a practice of utilizing multiple activation methods for peptide analysis so that paired spectra could be searched.[46-48] With ETD, ECD and CID being both widely available and at the same time often offering complementary information, it is not surprising that a majority of methods

that used paired data have focused on these activation methods. One of the first to do so was Savitski et al. who showed that using CID/ECD spectral pairs significantly improved identification of peptides via *de novo* sequencing.[44] A *de novo* sequencing program called Spectrum Fusion was developed by Datta and Bern which used the CID/ETD data and a Bayesian Network to create single combined spectra from the paired data.[46] An *et al.* demonstrated that paired CID/ETD spectra and on-tip derivatization of peptides by adding a fixed charge was successful for increasing the confidence of sequence identification.[47]

Described herein we have adapted our recently developed de novo sequencing method, UVnovo, to search CID/UVPD paired spectra, resulting in improved identification of peptides compared to that of CID or UVPD alone. In addition UVnovo has been modified to allow sequence tag de novo searches. Sequence tags searching is performed by *de novo* sequencing of only a small section of the peptide (typically 3 to 5 residues); these results are then searched against a database in a type of hybrid *de novo*/*in silico* search method. Using this method to search MS/MS data from *E.coli* lysate, the UVPD data resulted in a higher unique peptide count compared that to that of PEAKS and SEQUEST.

5.3 EXPERIMENTAL

5.3.1 Materials

Mass spectrometry grade trypsin-gold was purchased from Promega (Madison, WI, USA). LC-MS grade acetonitrile and water were purchased from EMD Millipore (Darmstadt, Germany). Phosphate buffered saline (PBS), dimethyl sulfoxide were

purchased from Thermo Fisher Scientific Inc. (San Jose, CA, USA)). Sulfosuccinimydyl-7-amino-4-methyl-coumarin-3-acetic acid (AMCA) was purchased from Pierce Biotechnology (Rockford, IL, USA). *E.coli* lysate was graciously donated by Dr. M. Stephen Trent research group at the University of Texas at Austin.

5.3.2 Modification *E.coli* lysate

Modification of E.coli lysate was performed as described previously.[45] In short *E.coli* lysate was carbamylated to block the reactive primary amines of the lysine side-chains by mixing 50 μg of the lysate in 50 mM sodium carbonate with 8 M urea and heating for 4 hours at 80 °C. The resulting carbamylated proteins were then buffer exchanged into PBS to remove urea and subsequently digested using trypsin at 37 °C overnight. After digestion, 25 μL of 20 mM AMCA in DMSO was added to the solution and kept in the dark overnight at room temperature. Following the AMCA modification a C18 SPE cartridge was used to clean the samples by removal of residual AMCA. Finally the samples were dried and reconstituted in an appropriate LC solvent.

5.3.3 LC-MS/MS analysis of *E.coli* lysate

Peptides were analyzed using a Dionex NSLC 3000 nanoLC system (Thermo Scientific; Waltham, MA, USA) interfaced to a Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific; San Jose, CA) that was previously modified for UVPD via a Coherent 351 nm excimer laser (Coherent; Santa Clara, CA, USA).[48] The laser was set to 3 mJ per pulse at 500 Hz, and 15 pulses were used for all experiments. Due to the complexity of the *E.coli* lysates, a 360 min gradient was used from 3% B increasing to 50% B with a flow rate of 300 nL/min. Mobile phase A was water with 0.1 % formic acid (v/v), and mobile phase B was acetonitrile with 0.1% formic acid (v/v)). Approximately 5 ug of peptide mixture was loaded on the column. The in-house packed

column was a 15 cm capillary packed with 3.5 µm particles (C18 stationary phase) with a pore size of 140 Å. Collision induced dissociation (NCE 35, 10 ms) and UVPD (15 pulses at 3 mJ per pulse) were undertaken for each sample. For the collection of paired CID/UVPD data, ion trap control language (ITCL) was customized to allow enhanced selectivity of activation parameters within the dual-pressure ion trap during LCMS analysis. Briefly, setting the NCE value to 0 for CID specified that isolated precursor ions were transferred to the low-pressure cell (LPC) followed by laser pulse triggering to obtain optimal UVPD results. In contrast, if the NCE parameter was set to a value above 0, the isolated precursor ions were retained in the high-pressure cell (HPC) for optimal CID fragmentation efficiency while laser pulse triggering would be inactive. After CID, fragment ions were transferred from the HPC to the LPC for subsequent detection.

5.3.4 PEAKS analysis

PEAKS 6 (Bioinformatics Solution Inc.; Waterloo, ON, Canada) was used as a comparison sequence tag search algorithm. Raw files were converted into mzXML and then imported into the PEAKS database. Standard ion trap search parameters were used for both de novo and database comparisons. The AMCA modification was added to the PTM section of the database as a fixed modification at the N-terminus. After completion of the PEAKS processing, the results were then checked against the Uniprot *E.coli* database.

5.3.5 SEQUEST analysis

In order to obtain a list of true peptides from the *E.coli* lysate, SEQUEST run through Proteome Discoverer v. 1.4 (ThermoFisher Scientific, San Jose, CA) was used.

AMCA was added as a fixed modification on the N-terminus of all peptides. SEQUEST was set up with standard settings except that lysine cleavage of trypsin was disabled by selecting Trypsin(R) in the settings. The database used was an *E. coli* reference database from UniProt.

5.3.6 UVnovo Modifications

A detailed description of UVnovo workflow has been described previously.[45] However new modification to the program were performed as follows

5.3.6.1 Merging UVPD/CID paired spectra

Processing and analysis of the paired CID/UVPD spectra proceeded in the same manner as described for the UVPD sample. The main difference, merging of spectral pairs into a single fragmentation probability spectrum, was accomplished by simply concatenating the predictor vectors into a single input for the random forest training and application. Again, the RF model was trained four times, with variable selection shrinking the number of predictors each iteration. The original set of 691 combined predictors from the UVPD and CID spectra was reduced to the set of the 33 most important variables, as determined automatically by the RF model. These were used to train a final RF model that could be immediately applied to paired CID/UVPD spectral data. All subsequent processing was identical to the previously described UVPD methods for generating and scoring de novo sequence reconstructions.

5.3.6.2 Hybrid sequence tag database search

We digest in silico the forward and shuffled E. coli protein coding sequences to enable a target/decoy search strategy. We additionally create a hash table mapping all kmers, for $3 \leq k \leq 6$, back to the the peptides they compose. This allows fast searching of de

novo sequence tags against the database for tags of length k. For each spectrum with a de novo sequence reconstruction, any database peptides with matching tags and a mass within ±2 Da of the precursor are scored against the sequence and HMM node scores. Oxidized methionine is allowed as an optional modification for this search.

We applied an optimization algorithm to a set of 4 derived peptide match features and trained a discriminative scoring function to maximize the number of target peptides identified at a 1% FDR.

5.4 RESULTS

5.4.1 UVnovo incorporation of paired spectra

E.coli lysate was digested and modified as described above. The carbamylation reaction blocked all primary amines at the protein level. Trypsin, inhibited from acting on the blocked lysines, cleaved C-terminal to arginine, generating peptides with new primary amines solely at the N-terminus. The subsequent AMCA reaction installed a UV chromophore at the N-terminus of each peptide, making the peptides well-suited for 351 nm UVPD.

ESI-MS/MS was performed using a dual cell linear ion trap mass spectrometer equipped with a 351 nm laser.[49] Following each MS1 survey scan, UVPD/CID spectra pairs were acquired for each of five selected precursors. The generation of UVPD/CID spectra required a modification of the instrument control software as CID is typically performed in the high pressure trap to ensure optimal dissociation and UVPD results are more optimal when carried out in the low pressure trap. Modification of the instrument

control software, as described in the experimental section, allowed seamless switching between UVPD and CID in alternating scans. The MS/MS spectra generated by UVPD for the AMCA-modified peptides exhibited series of *y* ions. Complementary *b* ions were annihilated during the multi-pulse UVPD period due to the high photoabsorptivity of the AMCA-tagged N-termini of the b ions; *y* ions did not contain a UV chromophore and survived as stable ions in the ion trap. The MS/MS spectra produced upon CID of the same AMCA-tagged peptides displayed conventional b and y ions with the b ions retaining the AMCA modification. **Figure 5.1** shows a set of paired UVPD/CID spectra for a peptide from Probable transcriptional regulatory protein YebC OS from the digested *E.coli* lysate.

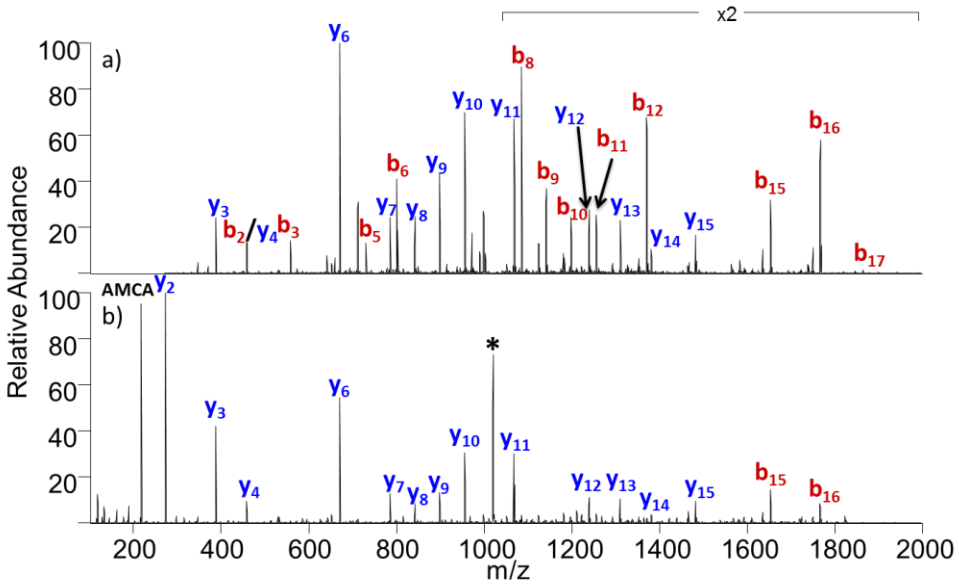


Figure 5.1: Example of MS/MS spectra observed for a single peptide when obtained via alternating CID and UVPD modes. The peptide shown is AMCA-tagged ELVTAAK(carbamyl)LGGGDPDANPR (2+) from an *E. coli* lysate digest. The precursor is labeled with an asterisk. a) CID (NCE 35) from scan 18750, and b) UVPD (3 mJ per pulse, 15 pulses) from scan 18751.

Spectra from an *E. coli* UVPD/CID paired run were processed using Proteome Discoverer SEQUEST with the Percolator node and allowing a ±1.8 Da precursor mass tolerance. Optional methioinine oxidation was included in addition to fixed N-terminal AMCA and carbamylated lysine modifications. Limiting the results to doubly charged precursors and top-ranked matches, 4,854 high confidence identifications matching 1957 unique peptides were obtained from a total of 51,525 spectral pairs (103,050 spectra) collected across all three replicates.

We processed the UVPD/CID spectral pair with our de novo analysis program, UVnovo, and benchmarked its performance against that obtained using only the CID or UVPD subset of scans from the same data set. Spectra, individual or paired, were transformed using a random forest classifier into a single vector of scores for each potential C-terminal fragmentation position. This output, along with empirically derived amino acid frequency statistics, was further refined into a series of fragmentation site probabilities using a hidden Markov model (HMM). Application of the Viterbi algorithm then found the most probable *de novo* sequences, and these were scored based on the individual and aggregate fragment probabilities determined by the HMM.

The paired UVPD/CID spectra were processed with UVnovo. The UVnovo results are displayed in **Figure 5.2** as the number of correct sequences binned according to peptide length and shown in comparison to the peptide spectral matches found by Sequest. The unfiltered data for the paired spectra returned 68.3% correct matches for top scored de novo sequences and 79.1% when the set was expanded to the top three

scored de novo matches for each spectrum (**Figure 5.2a** and **b**). As with the CID and UVPD data, low ranked de novo reconstructions were able to be removed which increased the results for the top ranked de novo match to 78.2% and the results for the top three ranked set to 88.1% (**Figure 5.2c** and **d**). In addition to the paired spectra both the

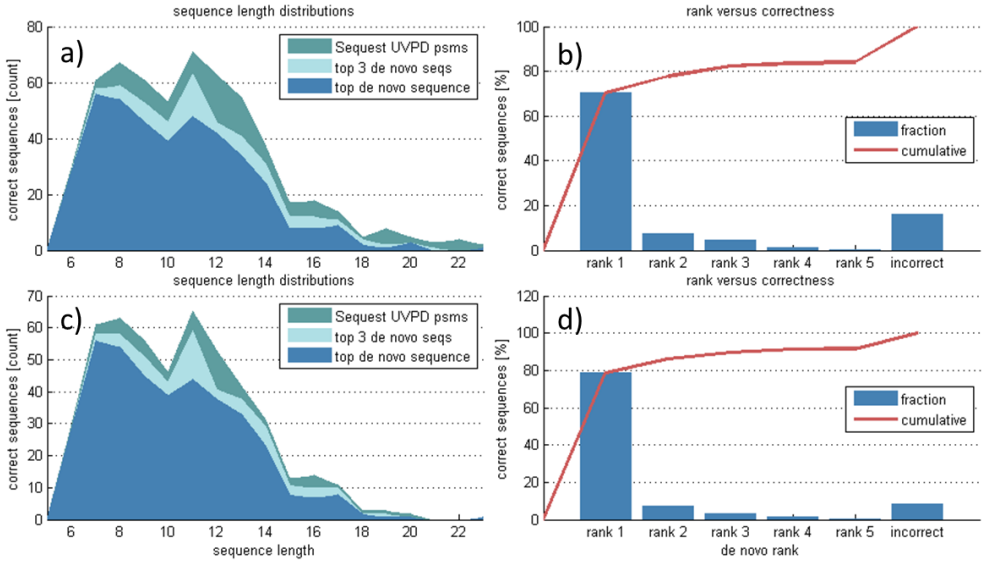


Figure 5.2: UVnovo *de novo* results for paired UVPD/CID spectra from *E.coli* lysate. No sequence gaps are allowed for a correct match. a) Graphical display of the number of peptides correctly identified (from SEQUEST results) grouped by peptide length in comparison to the number of correct matches based on the top *de novo* result and the number of correct matches based on the top three *de novo* results. b) Graphical display of the percentage of correct matches when considering all peptide lengths. The line shows the cumulative total upon addition of all correct matches from peptide sequences with a lower *de novo* rank. c) and d) Graphical displays showing the UVnovo results after the application of a filter to remove all low rank *de novo* reconstructions.

UVPD and CID spectra were processed using UVnovo separately in order to compare these to the paired spectra results. For UVPD, The top-ranked sequences generated were correct 62.9% and when considering the top three sequences constructed by UVnovo

73.9% of the returned sequences corresponded to an exact full length match to the SEQUEST PSM (**Figure 5.3a and b**). The application of a filter to remove low ranked *de novo* reconstructions resulted in an increase in the overall accuracy of the UVnovo results for both the top one and top three sets which were improved to 75.9% and 86.7%, respectively (**Figure 5.3c** and **d**). CID spectra were then searched using UVnovo which resulted in correct top ranked sequences for only 35.3% of the spectra; this value increased to 49.0% when considering the top three highest scored de novo sequences for each spectrum (**Figure 5.4a and b**). The CID results were also filtered in the same way as the UVPD, resulting in an increase in the results to 51.9% and 71.3% for the top one and top three de novo sequence sets, respectively (**Figure 5.4c and d**).

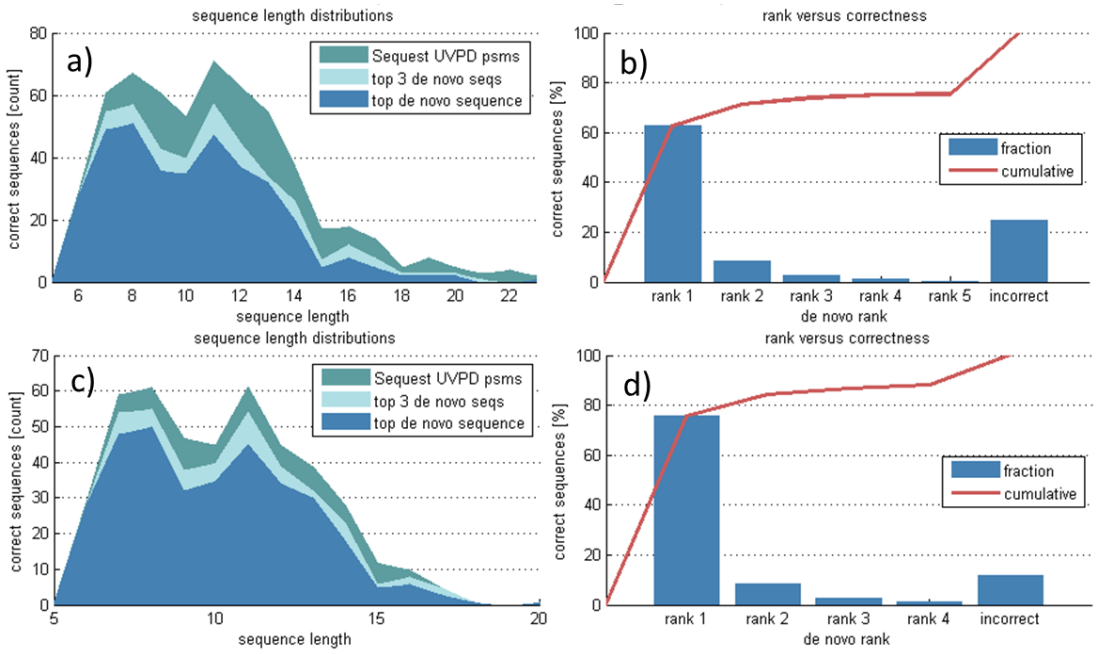


Figure 5.3: UVnovo *de novo* results for UVPD spectra from *E.coli* lysate. No sequence gaps are allowed for a correct match. a) Graphical display of the number of peptides correctly identified (from SEQUEST results) grouped by peptide length in comparison to the correct number of matches based on the top *de novo* result and the number of correct matches based on the top three *de novo* results. b) Graphical display of the percentage of correct matches when considering all sequence lengths. The line shows the cumulative total upon addition of correct matches from peptide sequences with a lower *de novo* rank. C) and d) Grapical displays showing the UVnovo results after the application of a filter to remove all low rank *de novo* reconstructions.

The relatively poor performance of CID is a direct result of the spectra containing both b and y type ions. The complementary nature of b and y type ions often results in mirror image ion ladders along the back bone from both N- and C-terminus. While redundant fragmentation can be useful if a cleavage site is in one the of the ion ladders to help fill in the gaps, it is also detrimental as it increases the difficulty of constructing a

sequence because the algorithm must distinguish between b- and y- fragments. As a result of this shortcoming the standalone UVPD outperforms CID due to the fact that *b* ions are eliminated from the MS/MS spectra.

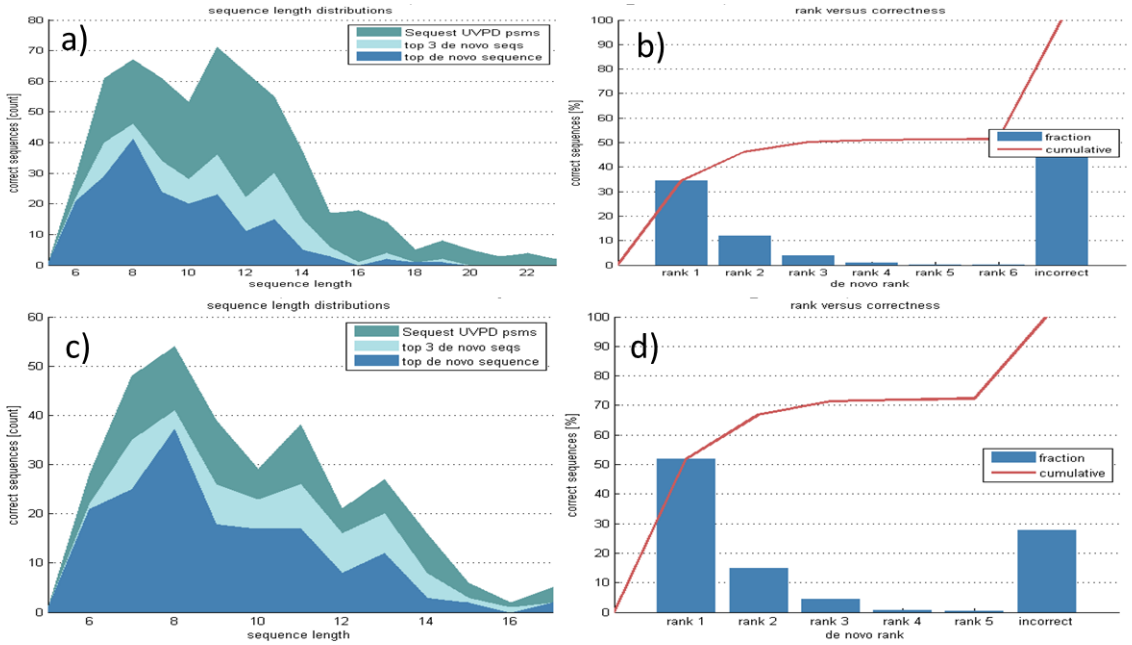


Figure 5.4: UVnovo *de novo* results for CID spectra from *E.coli* lysate. No sequence gaps are allowed for a correct match. a) Graphical display of the number of peptides correctly identified (from SEQUEST results) grouped by peptide length in comparison to the correct number of matches based on the top *de novo* result and the number of correct matches based on the top three *de novo* results. b) Graphical display of the percentage of correct matches when considering all sequence lengths. The line shows the cumulative total upon addition of correct matches from peptide sequences with a lower *de novo* rank. C) and d) Graphical displays showing the UVnovo results after the application of a filter to remove all low rank *de novo* reconstructions.

The combination of CID and UVPD spectra result in a noticeable improvement in the number of correct de novo reconstructions observed compared to either CID or UVPD alone. The paired UVPD/CID datasets capitalize on the benefits of both types of

spectra while ignoring or canceling out the drawbacks of either CID or UVPD spectra. As noted above, the redundant b/y ion ladders from CID impairs reconstruction of the correct sequences. UVPD does not suffer from this problem and allows more accurate de novo sequence reconstruction; however, the assignment of the precursor mass is more challenging from the UVPD spectra because complementary mirror image sequences provide more confidence with this task (a notable advantage of the CID spectra). Thus, combining the UVPD and CID spectra results in even more accurate results from UVnovo.

5.4.2 Sequence Tag Analysis

An additional modification of UVnovo allowed usage as a *de novo* tag based hybrid search algorithm. UVnovo was altered to facilitate incomplete sequence reconstructions, known as sequence tags. These tags are searched against a known database to generate candidate peptide sequences, which are then scored and ranked. The *E.coli* lysate data files were searched using UVnovo modified to perform sequence tag de novo sequencing and was searched against a Uniprot *E.coli* database with a shuffled decoy. Using this method, UVnovo matched 2,260 peptides identified from the *E.coli* lysate. UVnovo outperformed PEAKS, a *de novo* program that also uses sequence tag assignments, which identified 1,689 peptides from the same dataset. The enhanced performance of the custom algorithm was attributed to the absence of *b* ions in the spectra which normally complicate ion assignments in *de novo* workflows. The performance of UVnovo was also compared to SEQUEST, a traditional *in silico* database search program, which returned 1,957 peptide hits. As demonstrated for the *E coli* lysate, the reduction of spectral complexity obtained by AMCA-tagging and selective UVPD proved to be a strategic advantage for *de novo* sequencing.

5.5 CONCLUSIONS

Two unique modifications to the *de novo* program, UVnovo, have been implemented to capitalize on paired UVPD/CID spectra of AMCA-modified *E.coli* lysate tryptic peptides. UVnovo was used to search paired UVPD/CID spectra and compared to the analysis of UVPD or CID spectra separately. The UVPD/CID paired spectra out-performed both UVPD and CID as top ranked de novo reconstructions were correct for 68.3% of searched spectra (compared to 62.9% for UVPD alone and 35.3% for CID alone) and 79.1% when the top three scored sequences from UVnovo are considered (75.9% for UVPD alone and 51.9% for CID alone). These results are improved to 78.2% (compared to 75.9% for UVPD alone and 51.9% for CID alone) for the top result and 88.1% (86.7% for UVPD alone and 71.3% for CID alone) for the top three results when a filter is applied to remove low ranked *de novo* results. These results convey the successful use of UVnovo for increasing the confidence in the de novo sequencing of peptides using a paired CID/UVPD approach. Secondly UVnovo was altered to allow the generation of de novo sequence tags. These sequence tags were then searched against a database which resulted in the identification of 2,260 unique peptides from the *E.coli* lysate. Comparing this to another hybrid *de novo*/database search program (PEAKS) and a traditional in silico program (SEQUEST), UVnovo returned a larger number of unique peptides than either PEAKS (1689 peptides) or SEQUEST (1957 peptides). The UVnovo/paired UVPD/CID strategy improved the confidence of searches and enabled application alongside traditional database search algorithms.

5.6 REFERENCES

(1) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(2) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(3) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(4) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(5) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(6) De Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. Comprehensive Mass-Spectrometry-Based Proteome Quantification of Haploid versus Diploid Yeast. *Nature* **2008**, *455* (7217), 1251–1254.

(7) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(8) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(9) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(10) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(11) Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q. pNovo: De Novo Peptide Sequencing and Identification Using HCD Spectra. *J. Proteome Res.* **2010**, *9* (5), 2713–2724.

(12) Jeong, K.; Kim, S.; Pevzner, P. A. UniNovo: A Universal Tool for de Novo Peptide Sequencing. *Bioinformatics* **2013**, *29* (16), 1953–1962.

(13) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(14) Keough, T.; Lacey, M. P.; Youngquist, R. S. Solid-phase Derivatization of Tryptic Peptides for Rapid Protein Identification by Matrix-assisted Laser

Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16* (11), 1003–1015.

(15) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(16) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(17) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(18) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(19) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(20) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(21) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(22) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(23) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(24) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(25) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(26) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(27) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(28) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(29) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(30) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(31) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(32) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(33) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(34) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(35) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(36) Liu, X.; Shan, B.; Xin, L.; Ma, B. Better Score Function for Peptide Identification with ETD MS/MS Spectra. *BMC Bioinformatics* **2010**, *11* (Suppl 1), S4.

(37) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(38) Richards, A. L.; Vincent, C. E.; Guthals, A.; Rose, C. M.; Westphall, M. S.; Bandeira, N.; Coon, J. J. Neutron-Encoded Signatures Enable Product Ion Annotation from Tandem Mass Spectra. *Mol. Cell. Proteomics MCP* **2013**, *12* (12), 3812–3823.

(39) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(40) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(41) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(42) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(43) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(44) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. Proteomics-Grade de Novo Sequencing Approach. *J. Proteome Res.* **2005**, *4* (6), 2348–2354.

(45) Robotham, S. A.; Horton, A. P.; Cannon, J. R.; Cotham, V. C.; Marcotte, E. M.; Brodbelt, J. S. UVnovo: A Novel De Novo Sequencing Algorithm Using Single Series of Fragment Ions via Chromophore Tagging and 351 Nm UVPD. *Submitted*.

(46) Datta, R.; Bern, M. Spectrum Fusion: Using Multiple Mass Spectra for de Novo Peptide Sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **2009**, *16* (8), 1169–1182.

(47) An, M.; Zou, X.; Wang, Q.; Zhao, X.; Wu, J.; Xu, L.-M.; Shen, H.-Y.; Xiao, X.; He, D.; Ji, J. High-Confidence de Novo Peptide Sequencing Using Positive Charge Derivatization and Tandem MS Spectra Merging. *Anal. Chem.* **2013**, *85* (9), 4530–4537.

(48) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of

Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

Chapter 6

Comparison of Ultraviolet Photodissociation and Collision Induced Dissociation of Adrenocorticotropic Hormone Peptides

6.1 Overview

In an effort to better characterize the fragmentation pathways promoted by ultraviolet photoexcitation in comparison to collision induced dissociation (CID), six adrenocorticotropic hormone (ACTH) peptides in a range of charge states were subjected to 266 nm ultraviolet photodissociation (UVPD), 193 nm UVPD, and CID. Similar fragment ions and distributions were observed for 266 nm UVPD and 193 nm UVPD for all peptides investigated. While both UVPD and CID led to preferential cleavage of the Y-S bond for all ACTH peptides (except ACTH (1-39)), UVPD was far less dependent on charge state and location of basic sites for the production of C-terminal and N-terminal ions. For ACTH (1-16), ACTH (1-17), ACTH (1-24) and ACTH (1-39), changes in the distributions of fragment ion types (*a,b,c, x, y, z* and collectively N-terminal ions versus C-terminal ions) showed only minor changes upon UVPD for all charge states. In contrast, CID displayed significant changes in the fragment ion type distributions as a function of charge state, an outcome consistent with the dependence on the number and location of mobile protons that is not prominent for UVPD. Sequence coverages obtained by UVPD showed less dependence on charge state than those determined by CID, with the latter showing a consistent decrease in coverage as charge state increased.

6.2 Introduction

Significant advances in mass spectrometry for proteomics applications have evolved from the development of increasingly powerful informatics methods as well as

new ion activation techniques and more sophisticated MS/MS strategies to improve the characterization of peptides in complex mixtures. The latter include deployment of real-time decision tree methods[1,2], use of targeted ion monitoring methods[3–5], development of hybrid ion activation methods[6–8], and strategic ion manipulation based on ion-ion reactions[9,10] and innovative ion charging concepts[11,12]. There has also been growing effort to understand and optimize the fragmentation pathways of peptides to improve sensitivity, to enhance recognition of peptides in database searches, and in some cases to exploit preferential bond cleavages to provide greater specificity[13–15]. While collision- and electron-based methods remain the most universally popular activation methods[16–21], a number of alternatives (surfaced induced dissociation [22,23], photodissociation[24,25], high energy cation beam activation[26], metastable atom activation[27]) have been developed to afford higher energy deposition and expand the arsenal of ways to energize ions to create meaningful fragmentation patterns.

Photon-based activation methods (termed photodissociation) offer considerable versatility, and this has led to a number of applications for analysis of peptides and proteins. These methods include infrared multiphoton dissociation (IRMPD),[25,28–31] ultraviolet photodissociation (UVPD),[25,32–59] and visible photodissociation (Vis-PD).[25,60–62] The most common wavelengths used for UVPD of peptides include 157 nm, 193 nm, 266 nm, 351 nm and 355 nm, each corresponding to ones readily generated by pulsed excimer or YAG lasers. The photodissociation process requires that the peptides of interest possess a suitable chromophore, either via an intrinsic chromophore (e.g.,

amide bond or side-chains groups of amino acids) or ones added via derivatization of the peptides prior to analysis.[63–65] For example, peptides do not naturally absorb photons around 350 nm, but they can be tagged with appropriate chromophores to make them undergo photodissociation, ultimately producing conventional *b/y*-type fragment ions.[43-45] This strategy has been reported for the identification of antigen binding regions of antibodies[44], for streamlining bottom-up proteomics[45], and has also shown to useful in de novo sequencing by simplifying spectra.[43] The amide bond absorbs 157 and 193 nm photons, thus serving as the chromophore for 157 nm and 193 nm UVPD of peptides and proteins. The high energy per photon (7.9 eV for 157 nm and 6.4 eV for 193 nm) can elevate ions to excited electronic states, thus accounting for the production of a wide array of *a, b, c, x, y,* and *z* ions. Both 193 nm and 157 nm UVPD have been reported for numerous bottom-up proteomic applications and more recently for top down characterization of intact proteins.[36,37] A detailed investigation of 157 nm UVPD demonstrated that some of the fragmentation pathways were radical directed.[38] Also a recent study has shown that high energy fragmentation methods where electrons excited to almost ionized states precedes C-C backbone cleavages which lead to formation of a and x ions.[39]

In contrast to 193 nm and 157 nm photons, 266 nm photons are not absorbed by the peptide backbone but rather by the amino acid side chains of tyrosine, tryptophan and phenylalanine. Despite the lower energy per photon (4.6 eV), 266 nm UVPD also results in cleavages that lead to formation of *a,b,c,x,y* and *z* ions, similar to the types of ions

formed upon absorption of 157 nm or 193 nm photons.[25] Oh *et al.* reported the rich array of fragment ions upon 266 nm UVPD of protonated peptides and noted enhanced backbone cleavages adjacent to aromatic amino acids (tryptophan, phenylalanine and tyrosine).[47] The Kim group also showed that tryptic peptides derivatized by phenyl isothiocyanate reagents required lower laser powers for efficient photodissociation at 266 nm, thus confirming the enhanced photoabsorption cross-sections attributed to the added aromatic chromophores.[49] Kim and coworkers also analyzed phosphopeptides using 266 nm UVPD which led to the production of highly characteristic a_n – 97 ions affiliated with each phosphorylated residue.[57] Park *et al.* used 266 nm UVPD to analyze phosphorylated peptides in an FTICR mass spectrometer, finding a consistent loss of 98 Da.[50] The Julian group has shown that absorption of 266 nm photons promotes selective homolytic cleavage of disulfide and C-S bonds, thus providing a means to map cysteine residues in peptides and proteins.[51] Ly and Julian also utilized 266 nm photons to activate peptides or proteins containing iodinated tyrosines, leading to a process termed radical directed dissociation (RDD) which results in site-localized fragmentation and can be used as a spatially-specific probe of protein structure in the gas phase.[53,54,56] In another RDD strategy, Diedrich and Julian showed that phosphorylated sites in peptides could be pinpointed using a Michael addition reaction with naphthalenethiol followed by 266 nm UVPD to selectively cleave the C-S bond installed during the Michael addition reaction.[56] More recently Tao *et al.* showed that D- and L- amino acids in peptides could be differentiated using an RDD process initiated by 266 nm UVPD.[59]

This variety of photon-based activation methods has propelled interest in understanding the correlation between the photon wavelength and the outcomes of peptide activation in terms of dissociation efficiency, the types of fragmentation processes, and the potential for selective bond cleavages. Detailed comparisons of the fragmentation pathways and distributions of fragment ions arising from 157 nm, 193 nm, and 266 nm UVPD have not been extensively explored. Recently Lai *et al.* compared the fragmentation caused by 193 nm UVPD and 266 nm UVPD for singly protonated peptides, including angiotensin (DRVYIHPF) analogues and bradykinin (RPPGYSPFR).[66] They noted numerous similarities in the types of fragment ions (both ones derived from amide and non-amide backbone cleavages and some side-chain cleavages) produced by 193 nm and 266 nm UVPD along with some variations in relative abundances.[66] Herein we report the UVPD patterns of six adrenocorticotropic hormone (ACTH) peptides as a function of charge state and amino acid composition. ACTH peptides were chosen as they contain the three amino acids with aromatic side-chains (Phe, Trp, Tyr) that serve as the optimal chromophores for absorption of 266 nm photons (as well as secondary chromophores for 193 nm photons). Our goal is to extend the fundamental understanding of 266 nm UVPD in comparison to 193 nm UVPD and collisional activation.

6.3 EXPERIMENTAL

6.3.1 Materials

The six ACTH (human) peptides, ACTH (1-10, sequence SYSMEHFRWG, M_r 1298.4), ACTH (1-14, sequence SYSMEHFRWGKPVG, M_r 1679.9), ACTH (1-16, sequence SYSMEHFRWGKPVGKK, M_r 1936.3), ACTH (1-17, sequence SYSMEHFRWGKPVGKKR, M_r 2092.4), ACTH (1-24, sequence SYSMEHFRWGKPVGKKRRPVKVYP, M_r 2932.5), ACTH (1-39, sequence SYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAEAFPLEF, M_r 4540.1), as well as Alpha-Bag Cell peptide (1-9 sequence APRLRFYSL M_r 1122.3), [Tyr5] Bradykinin (sequence RPPGYSPFR M_r 1076.2), [Phe2, Nle4] ACTH (1-24 sequence SFS(Nle)EHFRWGKPVGKKRRPVKVYP, M_r 2899.5), and Adamtsostatin-16 (sequence SPWSQCATSCGGGVQTR with a disulfide between the two cysteiens M_r 1722.9) were purchased from American Peptide Company (Sunnyvale, CA, USA). LC-MS grade acetonitrile and water were obtained from EMD Millipore (Darmstadt, Germany). LC-MS grade formic acid was purchased from Fisher scientific (Fair Lawn, NJ).

6.3.2 Mass Spectrometry

All experiments were carried out using a Thermo Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific; San Jose, CA) or a Thermo Orbitrap Elite mass spectrometer (Thermo Scientific; Bremen, Germany). The dual linear ion trap was used for the MS/MS comparisons based on 266 nm UVPD, 193 nm UVPD, or CID, and the orbitrap instrument was used primarily to confirm specific fragment ion assignments based on high resolution, high accuracy measurements. The dual linear ion trap was modified to allow UVPD with a Continuum Minilite Nd:YAG laser (Santa Clara, CA, USA) set to 266 nm or a Coherent Excistar excimer laser (Santa Clara, CA) set to 193

nm. The set-up and implementation of UVPD was similar to that described previously.[31] Peptides were diluted in a 50/50 acetonitrile/water solution with 0.1% formic acid to a final concentration of 1 μM and were infused at a flow rate of 3 μL/min. CID was performed on every observed charged state of each of the six ACTH peptides using a normalized collision energy of 35%. 266 nm UVPD was performed on each charge state observed for the six ACTH peptides using 1, 2, 5, or 10 pulses (nominally 6 mJ per pulse) at a 10 Hz laser pulse rate. 193 nm UVPD was also performed for all observed charge states of the six ACTH peptides using 2 pulses and 2 mJ per pulse at 500 Hz laser pulse rate. Only a fraction of the light enters the ion trap due to the fact that neither focusing nor collimating optics are used, and the laser beam is divergent as it emerges from the laser. Due to the relatively low photon flux through the ion trap, the extent of secondary (or consecutive) fragment ion dissociation is minimal.

6.3.3 Analysis of Spectra

All CID and UVPD spectra were analyzed manually. Lists of fragments were obtained using both Protein Prospector v. 5.12.2 (UCSF) and Proteomics Tool kit (Institute for Systems Biology). Data for the fragment ion distributions was processed using Microsoft Excel. Abundance information was extracted from the raw data and used to calculate the percentage distributions for each type (*b* and *y* for CID and *a, b, c, x, y,* and *z* for UVPD).

6.4 RESULTS

6.4.1 266nm UVPD, 193 nm UPVD, and CID of ACTH peptides

UVPD using 266 nm photons has been less frequently employed for analysis of peptides than 193 nm UVPD, and it seemed particularly important to evaluate the

fragmentation induced by 266 nm photon absorption as well as the similarities and differences to that caused by 193 nm photon absorption. In order to investigate the variations in peptide fragmentation obtained by 266 nm UVPD, each charge state of the six ACTH peptides was activated using one to ten laser pulses. MS/MS spectra were also acquired by conventional CID and by 193 nm UVPD (2 pulses) to compare the types and distribution of fragment ions obtained by each of the three activation methods. Examples of the MS/MS spectra produced by 266 nm UVPD (5 laser pulses), 193 nm UVPD, and CID are shown in **Figure 6.1** for ACTH (1-17) (3+ charge state).

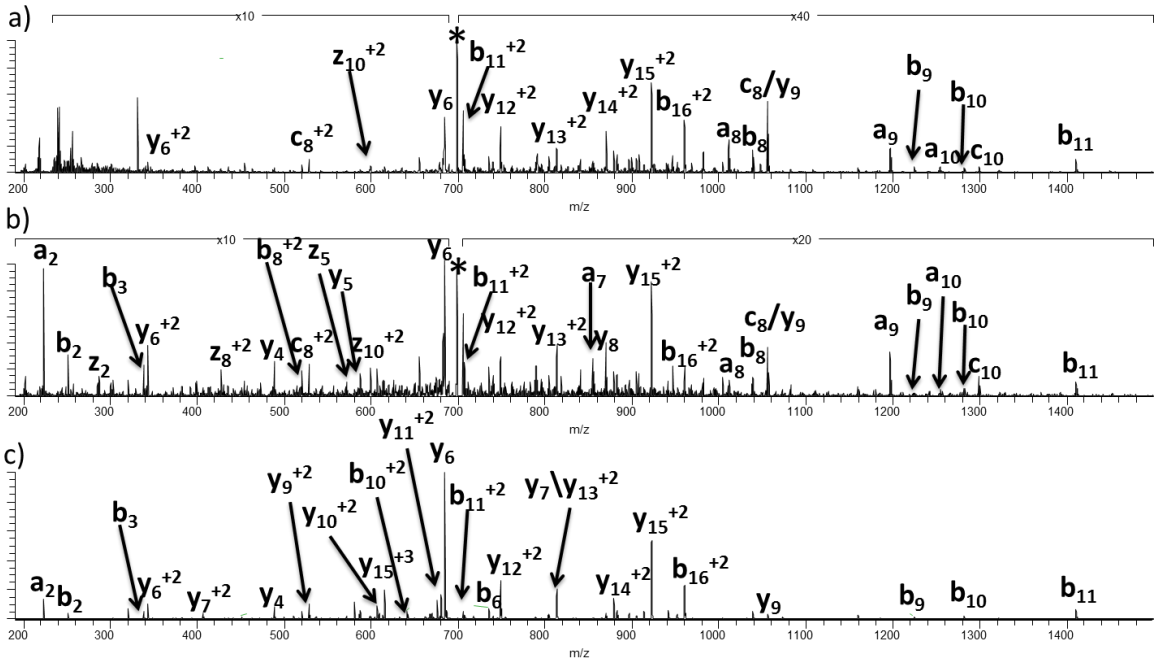


Figure 6.1: MS/MS spectra of 266 nm UVPD, 193 nm UVPD, and CID fragmentation of the ACTH (1-17). Precursor Ion is in the 3+ charge state labeled with an asterisk. . a) 266 nm UVPD (5 pulses 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

Both the 266 nm UVPD and 193 nm UVPD spectra show a rich array of fragment ions, ranging from the conventional *a/b/y* ions observed upon CID to *c/z* ions typically observed upon electron-activated (radical type) dissociation and *x* ions that are only commonly seen upon UVPD. The fragment ions in the visually congested UVPD spectra can be confidently assigned based on high accuracy measurements. **Figure 6.2** shows an expansion of the region from m/z 760 to 890 for peptide ACTH (1-24) (4+).

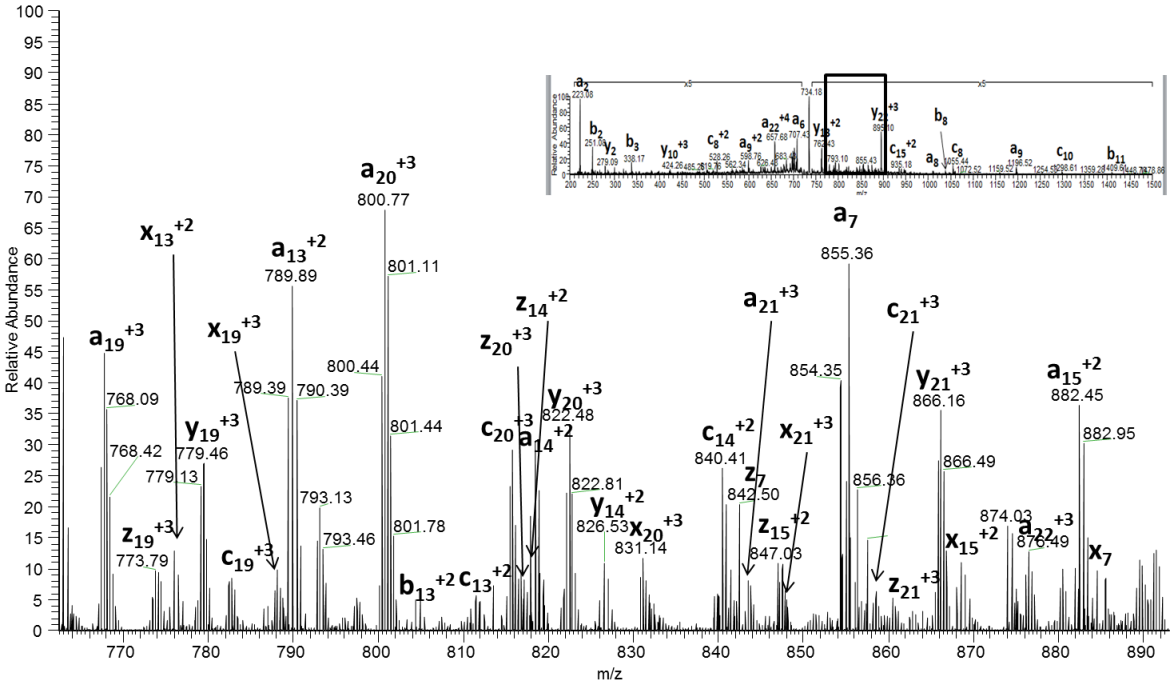
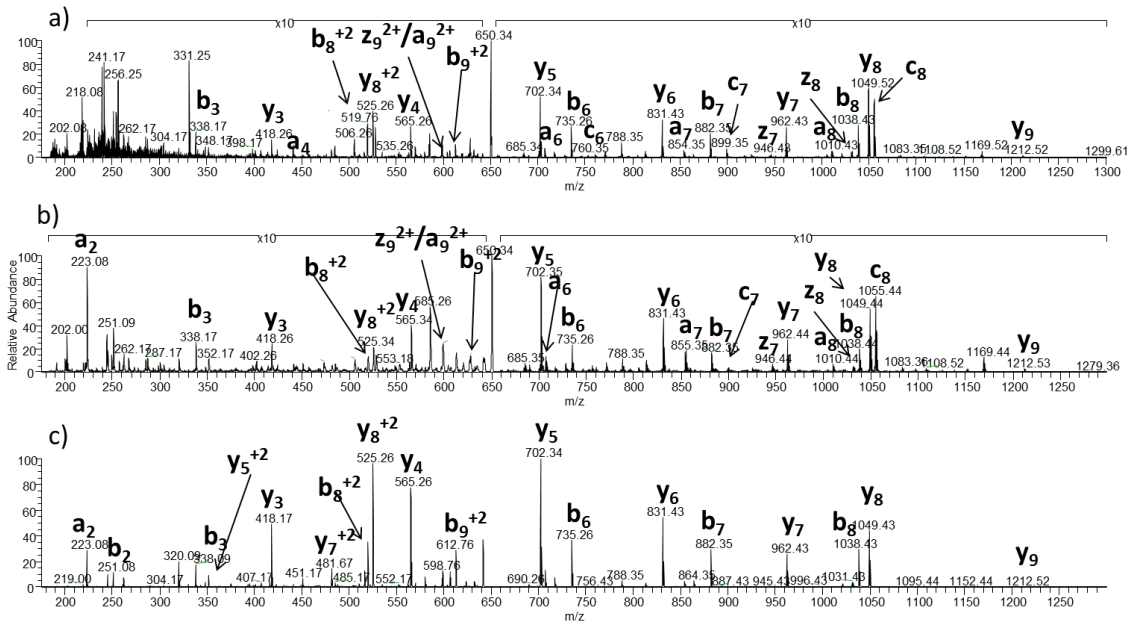


Figure 6.2: Expanded region of the UVPD mass spectrum from m/z 760 to 890 for ACTH (1-24) 4+ spectra showing the rich fragmentation observed upon UVPD. The spectrum was collected on a high mass accuracy Thermo Orbitrap Elite mass spectrometer to ensure correct identification of highly charged fragment ions.

UVPD (both 266 nm and 193 nm) and CID mass spectra for the other ACTH peptides are shown in **Figures 6.3-7.**



Supplemental Figure 6.3: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-10). Precursor ion is in the 2+ charge state. a) 266 nm UVPD (5 pulses; 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

Histograms summarizing the distribution of fragment ions obtained by 266 nm UVPD, 193 nm UVPD, and CID are shown in **Figure 6.8** for ACTH (1-17) for each of its five charge states (2+, 3+, 4+, 5+, 6+), and additional histograms for all of the MS/MS data for each activation method are shown in **Figures 6.9** (266 nm UVPD), **6.10** (193 nm UVPD) and **6.11** (CID). The most prominent fragment ions produced by each activation method for each charge state of each peptide are summarized in **Table 6.1**.

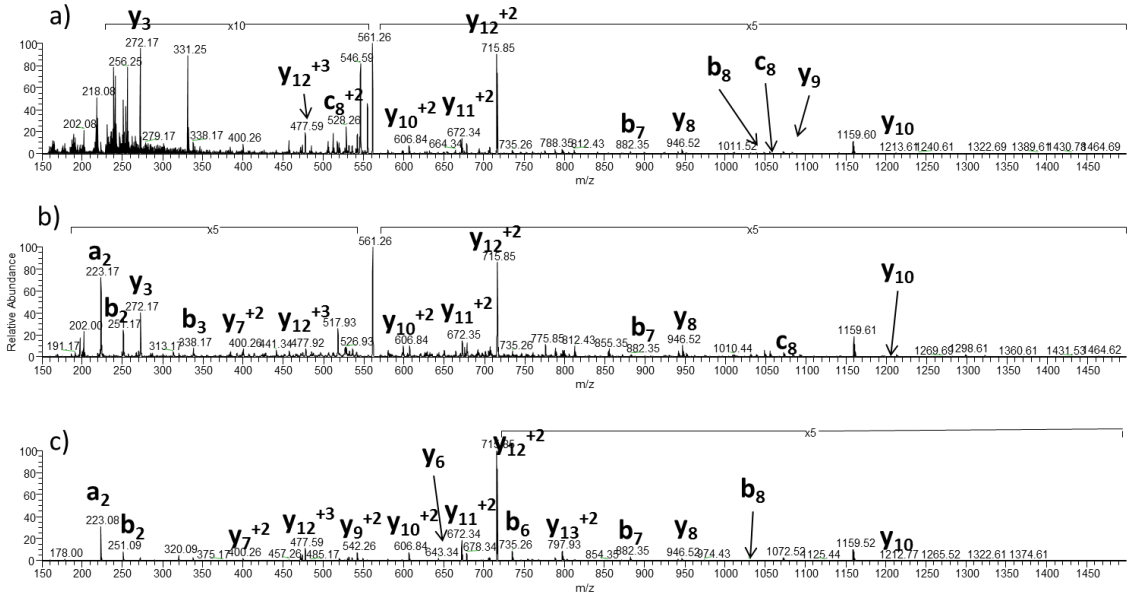


Figure 6.4: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-14). Precursor ion is in the 3+ charge state. a) 266 nm UVPD (5 pulses, 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

As expected, the number of laser pulses influenced the total abundance of fragment ions produced upon 266 nm UVPD as well as the distribution of fragment ions to a lesser extent. **Figure 6.12** shows the increase in the total fragment abundance for each charge state observed for the ACTH peptides as the number of laser pulses increases. The total abundances of fragment ions generally increased with the number of pulses, a result expected because the precursor ion population was not annihilated using a single laser pulse and thus more precursor ions were converted to fragment ions upon each subsequent pulse. The number of laser pulses used for 266 nm UVPD had a relatively minor influence on the distribution of fragment ions as

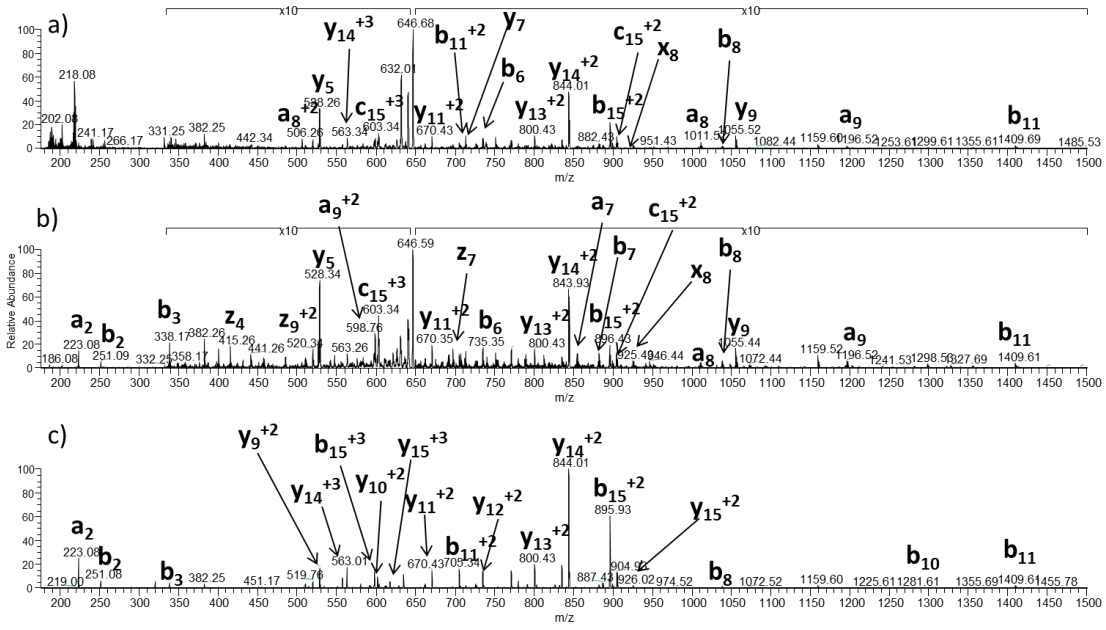


Figure 6.5: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-16). Precursor ion is in the 3+ charge state. a) 266 nm UVPD (5 pulses 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

exemplified by the spectra shown in **Figure 6.13** for ACTH (1-17) and summarized in bar graph format in **Figure 6.14** for the highest and lowest charge states of each peptide. For the bar graph format, the abundances of a, b, c, x, y, and z fragment ions were summed collectively based on ion type as a function of the number of laser pulses.

Based on the graphical distributions shown in **Figures 6.12** and **6.14** and the representative series of spectra in **Figure 6.13,** it is clear that the number of laser pulses has a relatively modest impact on the overall distribution of fragment ions for our UVPD set-up. Visual comparison of the spectra in **Figure 6.13**

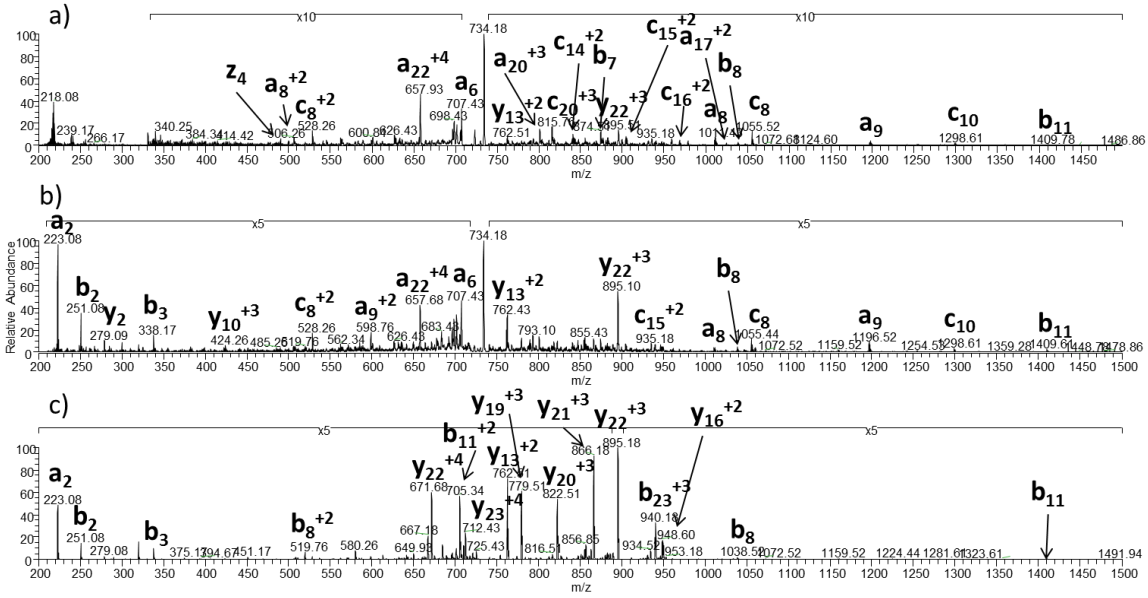


Figure 6.6: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-24). Precursor ion is in the 4+ charge state. a) 266 nm UVPD (5 pulses, 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

confirms that no new abundant ions are produced with an increasing number of laser pulses, suggesting that the majority of fragment ions are produced directly from the precursor ion, not via secondary dissociation of primary fragment ions. (This outcome depends on the laser power and the overlap of the laser beam with the ion cloud, so it would be expected to be instrument-dependent). For all spectral comparisons discussed in the next section, 5 laser pulses were used for the 266 nm UVPD spectra and 2 laser pulses were used for the 193 nm UVPD spectra.

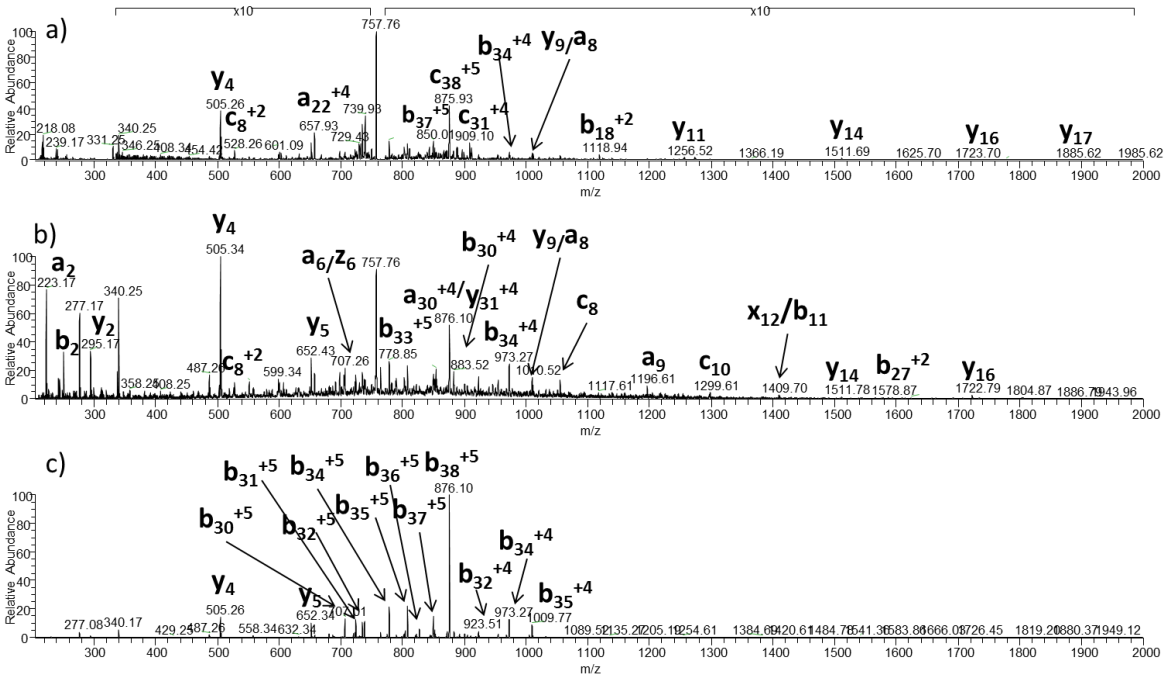


Figure 6.7: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-39). Precursor ion is in the 6+ charge state. a) 266 nm UVPD (5 pulses, 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

6.4.2 Fragmentation Trends Observed in ACTH peptides

The MS/MS spectra shown for ACTH (1-17) in **Figure 6.1** and the stacked histograms in **Figure 6.8** highlight some of the characteristic features observed for each activation method. The distributions of N-terminal (*a,b,c*) versus C-terminal ions (*x,y,z*) ions produced by 266 nm UVPD versus 193 nm UVPD are remarkably similar for each ACTH peptide in every charge state. For example, for the longer peptides (e.g. ACTH (1-24) and ACTH (1-39)), the portion of C-terminal ions formed upon 266 nm UVPD or

193 nm UVPD remains around 50-55%, regardless of charge state, and the histogram

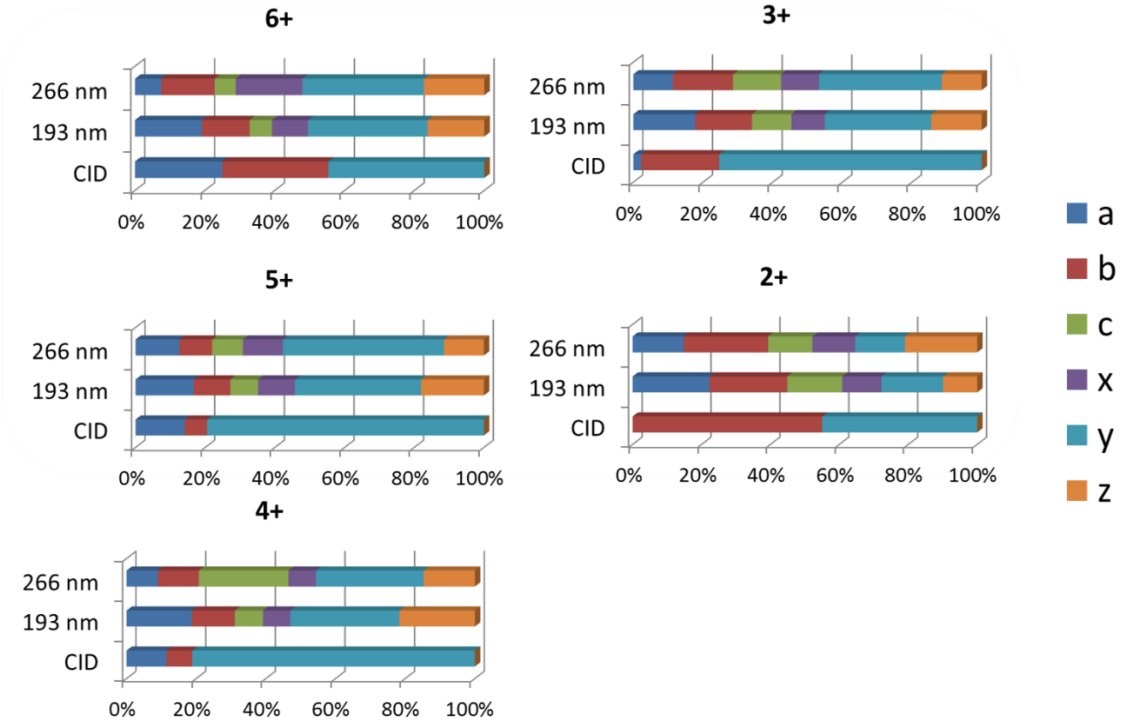


Figure 6.8: Distribution of 266 nm UVPD, 193 nm UVPD, and CID fragment ions for all observed charge states of ACTH (1-17). 5 pulses at 6 mJ was used for 266 nm UVPD, 2 pulses at 2 mJ was used for 193 nm UVPD, and 35% NCE was used for CID

graphs in **Figure 6.9** (for 266 nm UVPD) and **Figure 6.10** (for 193 nm UVPD) mirror each other for each peptide. The distribution of C-terminal versus N-terminal ions changes more dramatically for the shorter ACTH peptides (such as ACTH (1-10) and ACTH (1-14)), for which the portion of C-terminal product ions consistently increases from approximately 30% for the singly-charged peptides to upwards of 75% for the 3+ charge state. An increase in the portion of C-terminal ions with increasing charge state is

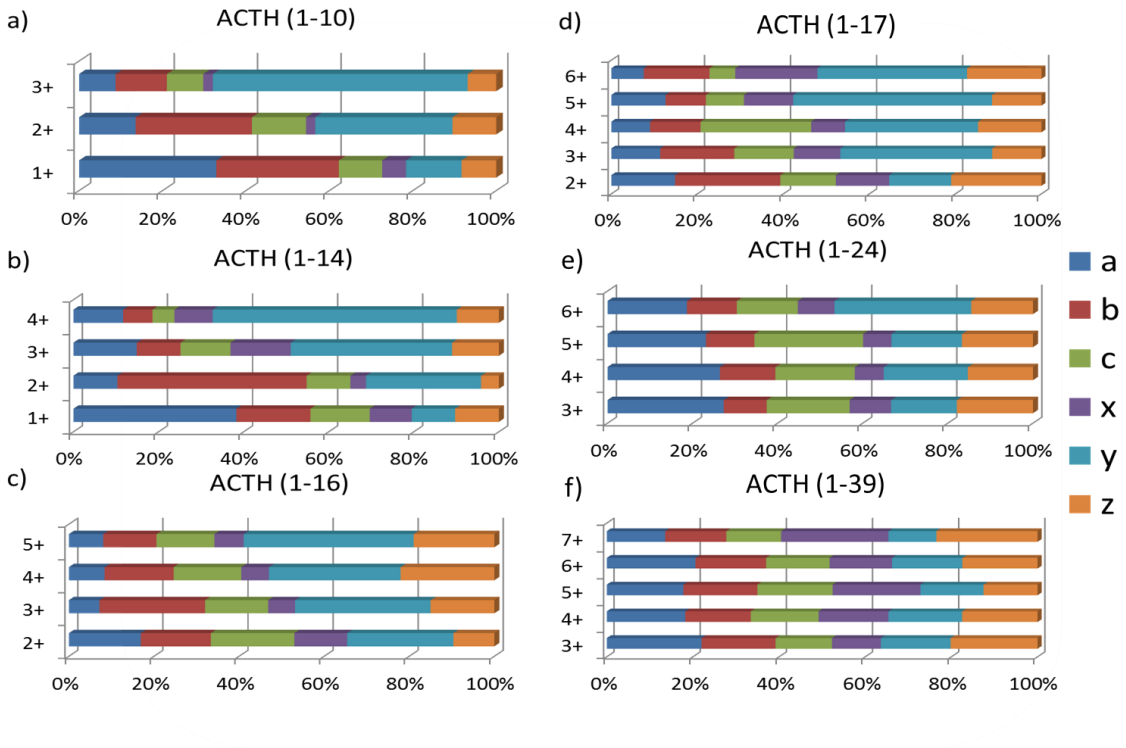


Figure 6.9: Distribution of 266 nm UVPD fragment ions for all observed charge states of a) ACTH (1-10), b) ACTH (1-14), c) ACTH (1-16), d) ACTH (1-17), e) ACTH (1-24), and f) ACTH (1-39). UVPD was carried out using 5 pulses at 6 mJ per pulse.

similarly observed for ACTH (1-10), ACTH (1-14), ACTH (1-16) and ACTH (1-17) for both UVPD methods. This increase in the portion of C-terminal (*x,y,z*) ions with increasing charge state is likely attributed to the predominance of basic sites on the second half of the peptide sequence, thus increasing the probability of charge retention for the C-terminal ions, but this effect does not necessarily reflect charge-modulated fragmentation directed by mobile protons as anticipated for CID. For the larger ACTH

ACTH (1-10)

Charge State	266 nm	193 nm	CID
3+	y_9^{+2} , y_5, y_6, y_8^{+2}	y_8^{+2}, a_2, y_4	y_8^{+2}, b_2, a_2
2+	c_8, y_8, b_8, y_5	a_2, y_5 , c_8	y_5, y_8^{+2} , y_4
1+	a_8, b_8, b_7	a_9, z_9, a_8, y_5	y_5 , b_8, b_7, b_6

ACTH (1-14)

Charge State	266 nm	193 nm	CID
4+	y_{12}^{+3}, y_{13}^{+3}, y_{10}^{+2}	y_{12}^{+3}, a_2, c_6^{+2}	y_{12}^{+3}, a_2, y_{13}^{+3}
3+	y_{12}^{+2} , y_{11}^{+2}, y_{10}^{+2}	a_2, y_{12}^{+2} , y_3	y_{12}^{+2} , a_2, y_{11}^{+2}
2+	b_{11}, b_{13}^{+2} , c_8, y_9	y_3, a_9, c_8, b_{11}	b_{11}^{+2} , b_{13}^{+2} , y_9, y_8
1+	a_9, z_{10}, c_8, a_8	a_9, y_9, b_8, a_8	y_9, b_{11}, b_8

ACTH (1-16)

Charge State	266 nm	193 nm	CID
5+	y_{14}^{+4} , y_{12}^{+3} , c_8^{+2}	y_{14}^{+4}, a_2, b_2, y_5	y_{14}^{+4}, a_2, y_{15}^{+4}
4+	y_{14}^{+3} , c_{15}^{+3} , y_5	y_{14}^{+3}, a_2, b_2	y_{14}^{+3}, a_2, b_{15}^{+3}
3+	y_{14}^{+2} , b_{15}^{+2} , c_{15}^{+2}	a_2, y_5 , y_{14}^{+2}	y_{14}^{+2} , b_{15}^{+2} , a_2
2+	y_9, a_9 , b_{15}^{+2}	a_9, y_5, b_{11}	b_{15}^{+2} , y_8 , b_{11}

ACTH (1-17)

Charge State	266 nm	193 nm	CID
6+	y_{15}^{+5}, a_{11}^{+3} , y_{15}^{+4} ,y_{14}^{+4}	a_2, y_{15}^{+5}, y_4	y_{15}^{+5} , a_2, y_7^{+2} , y_{13}^{+4}
5+	y_{15}^{+4}, y_{16}^{+4} , y_{13}^{+3}	y_{15}^{+4}, a_2, y_6, y_{13}^{+3}	y_{15}^{+4} , a_2, y_{16}^{+4}
4+	y_{15}^{+3}, y_6 , c_5	y_{15}^{+3}, a_2, y_6	y_{15}^{+3} , a_2, y_{14}^{+3}
3+	y_6 , y_{15}^{+2} , b_{16}^{+2}	a_2, y_6 , y_{15}^{+2}	y_6, y_{15}^{+2} , b_{16}^{+2} , y_{12}^{+2}
2+	a_9, b_{16}^{2+}, b_8, a_8, y_8	a_9, b_{16}, c_{10}, a_{13}	b_{16}, y_{12}^{2+}, b_{11}, y_9

ACTH (1-24)

Charge State	266 nm	193 nm	CID
6+	y_{22}^{+5}, b_{13}^{+3}, b_{20}^{+4}	a_2, y_{22}^{+5}, b_2, y_{13}^{+3}	y_{22}^{+5}, a_2, b_{23}^{+5}
5+	a_{22}^{+4}, y_{22}^{+4}, c_{20}^{+4}, a_8^{+2}	a_2, y_{22}^{+4}, b_2, y_{13}^{+2}	y_{22}^{+4}, b_{23}^{+4}, a_2
4+	a_{22}^{+4}, c_{20}^{+3}, a_{20}^{+3}, y_{22}^{+3}	a_2, y_{22}^{+3}, a_6, a_{22}^{+4}	y_{22}^{+3}, y_{21}^{+3}, y_{13}^{+2}
3+	a_{22}^{+3}, a_{23}^{+3}, c_8	a_{22}^{+3}, a_9, a_{20}^{+3}	y_{19}^{+3}, b_{11}, b_8

ACTH (1-39)

Charge State	266 nm	193 nm	CID
7+	b_{33}^{+5}, b_{34}^{+5}, y_4	a_2, y_4, b_2	b_{33}^{+5}, b_{31}^{+5}, b_{32}^{+5}
6+	b_{38}^{+5}, y_4, a_{22}^{+4}	y_4, a_2, b_{38}^{+5}	b_{38}^{+5}, b_{35}^{+5}, b_{34}^{+5}
5+	y_4, b_{35}^{+4}, a_{22}^{+4}, z_{25}^{+3}	y_4, z_{26}^{+3}, b_{25}^{+3}	b_{35}^{+4}, b_{34}^{+4}, y_4, b_{38}^{+4}
4+	b_{35}^{+3}, y_4, b_{19}^{+2}, y_{31}^{+4}	y_4, b_{35}^{+3}, a_9	b_{35}^{+3}, b_{38}^{+4}, b_{34}^{+3}
3+	c_8, y_{30}^{+2}, z_{31}^{+2}	y_4, c_{22}^{+2}, a_9	b_{29}^{+2}, b_{33}^{+3}, b_{28}^{+2}

[1]Ions are listed from most abundant to least abundant

Table 6.1: Comparison of Most Abundant Fragment Ions for ACTH Peptides[1]

peptides, fluctuations in the distribution of N-termini versus C-termini ions upon UVPD are less notable, with variations of less than 10% for most of the ion types (*a,b,c,x,y,z*).

The distributions of fragment ions produced by CID are shown in the stacked histograms of **Figure 6.11.** The CID results are much more dependent on the initial charge state of the precursor peptide, with swings as much as 50% for C-terminal versus N-terminal ion types (e.g. 32% C-terminal ions for doubly-protonated ACTH (1-14) versus 82% C-terminal ions for triply-protonated ACTH (1-14)) and distributions that

seesaw dramatically depending on the precursor charge state (e.g. 65% C-terminal ions for singly

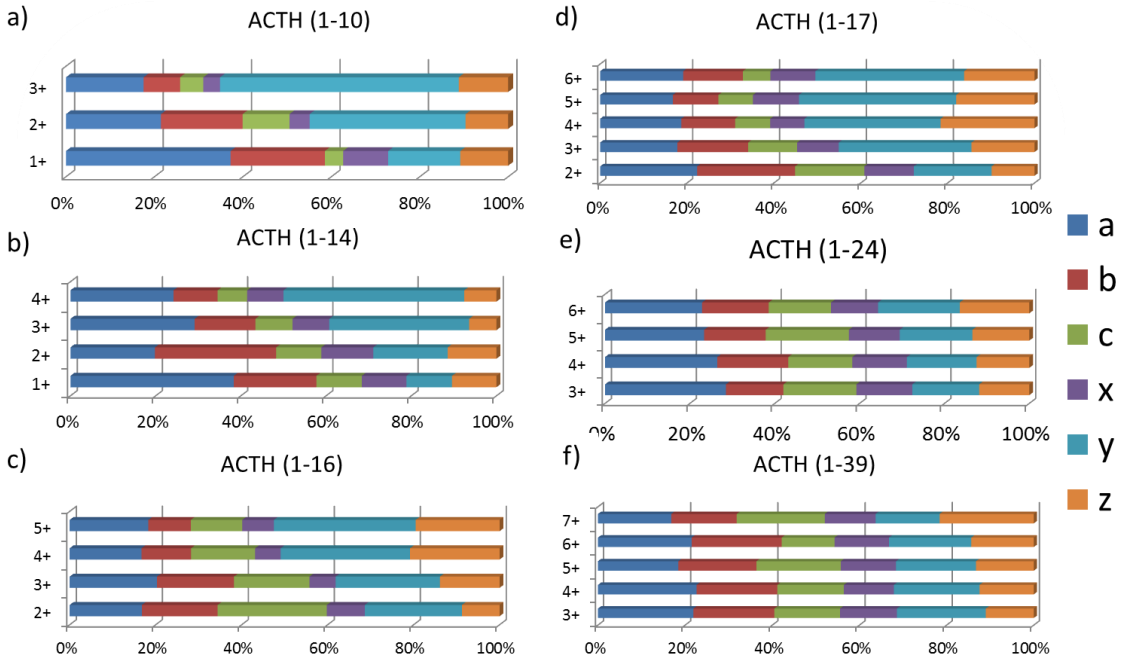


Figure 6.10: Distribution of 193 nm UVPD fragment ions for all observed charge states of a) ACTH (1-10), b) ACTH (1-14), c) ACTH (1-16), d) ACTH (1-17), e) ACTH (1-24), and f) ACTH (1-39). 193 nm UVPD was carried out using 2 pulses at 2 mJ per pulse.

protonated ACTH (1-14), 32% C-terminal ions for doubly-protonated ACTH (1-14), and 82% C-terminal ions for triply-protonated ACTH (1-14)). The shifts and seesaws in the C-/N-terminal fragment ion distributions for the CID data as a function of precursor charge state are not surprising. CID pathways are known to be highly charge-mediated

with the number of mobile protons (i.e. ones not sequestered at highly basic sites) dictating the preferred fragmentation processes.[67] The lack of significant variation in the

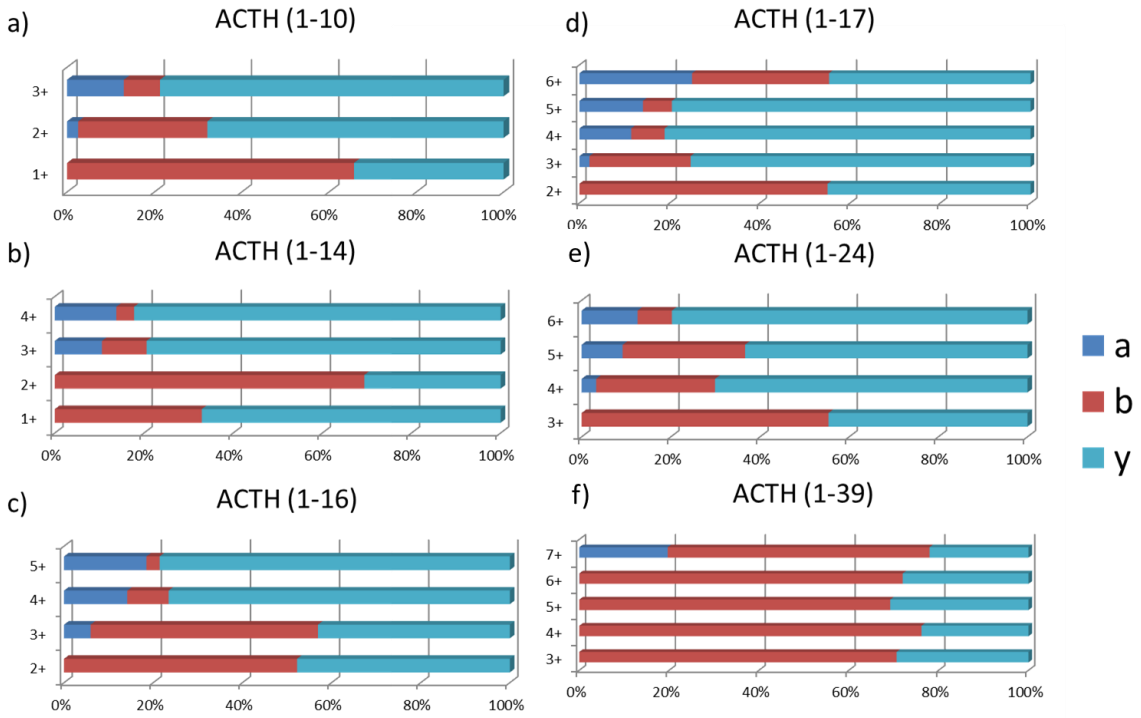


Figure 6.11: Distribution of CID fragment ions for all observed charge states of a) ACTH (1-10), b) ACTH (1-14), c) ACTH (1-16), d) ACTH (1-17), e) ACTH (1-24), and f) ACTH (1-39). CID was carried out using NCE of 35%.

C-terminal/N-terminal fragment ion distributions for either 266 nm UVPD or 193 nm UVPD reflects the reduced impact of proton-mediated pathways and the greater contribution from dissociation processes that may occur directly from excited states and/or radical-modulated ones. In addition, the rich diversity of UVPD fragment ions, as evidenced by the multi-segmented bars in **Figures 6.8 to 6.10,** underscores the wide

variety of competing (and possibly consecutive) pathways that minimize substantial variations as a function of charge state.

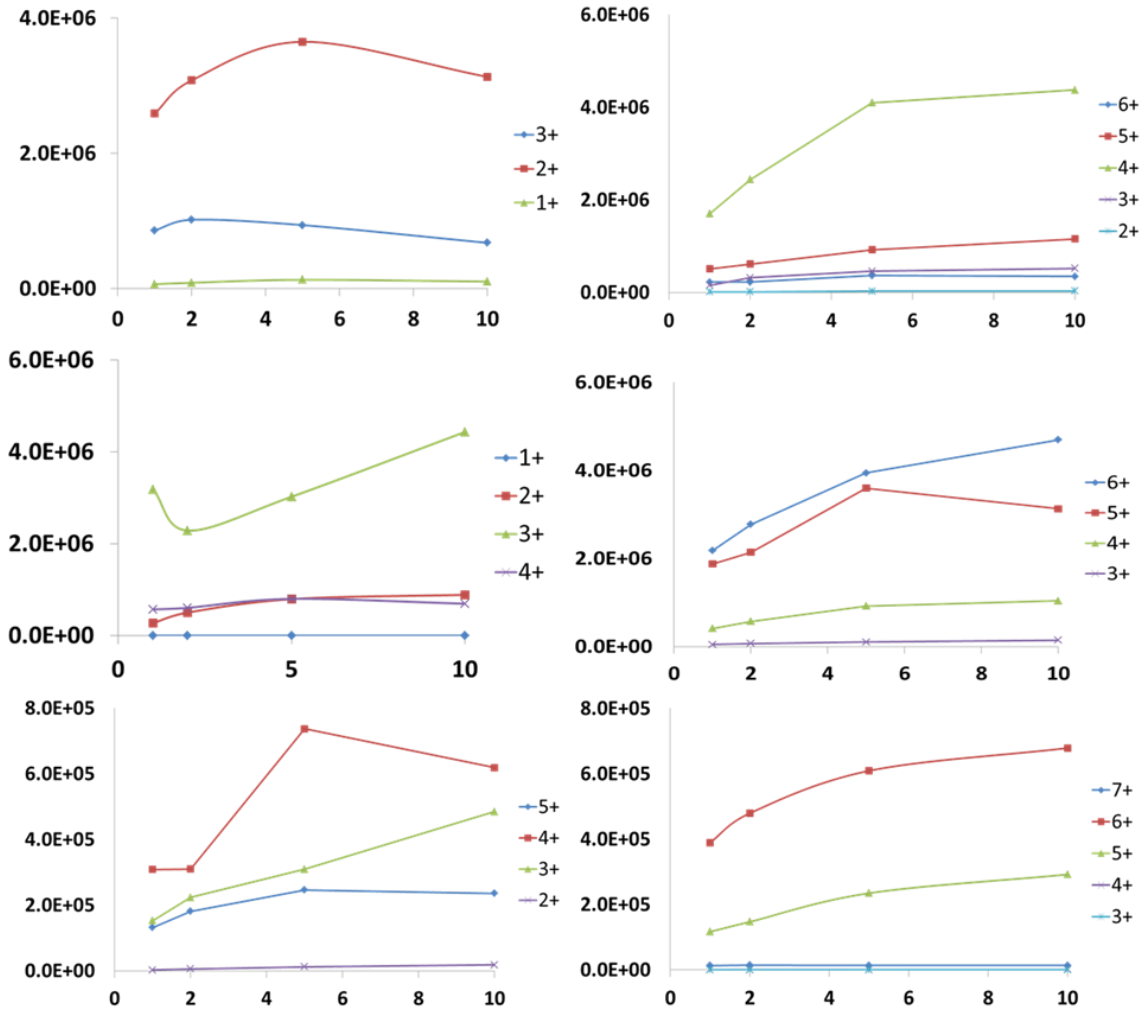


Figure 6.12: Total intensity of fragment ions generated by 266 nm UVPD for all charge states using 1, 2, 5, or 10 pulses for a) ACTH (1-10), b) ACTH (1-14), c) ACTH (1-16), d) ACTH (1-17), e) ACTH (1-24), and f) ACTH (1-39). The total abundances of fragment ions generally increased with the number of pulses, a result expected because the precursor ion population was not annihilated using a single laser pulse and thus more precursor ions were converted to fragment ions upon each subsequent pulse.

The MS/MS spectra for the ACTH peptides are frequently dominated by several highly abundant fragment ions which are indicative of preferential cleavages. The most

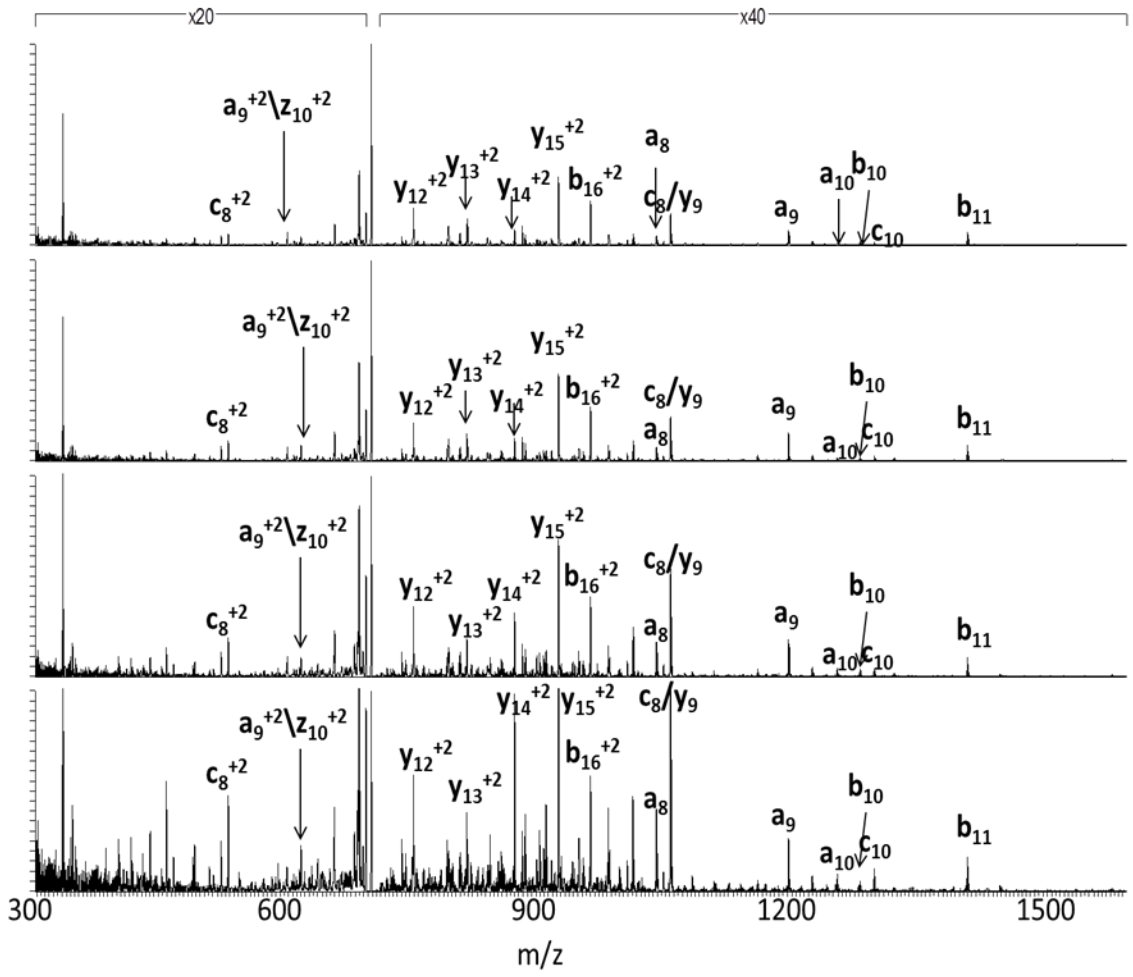


Figure 6.13: 266 nm UVPD mass spectra of ACTH (1-17), 3+ charge state, activated using a varying number of laser pulses. a) 1 pulse, b) 2 pulses, c) 5 pulses, and d) 10 pulses at 6 mJ per pulse. The number of laser pulses used for 266 nm UVPD had a relatively minor influence on the distribution of fragment ions.

predominant fragment ions produced by UVPD and CID are compiled in **Table 6.1** for each charge state of each ACTH peptide. For ACTH (1-10), several of the most dominant

fragment ions, namely a_2, y_4, y_5, and y_8, are observed for all three activation methods, with cleavage of the backbone at the Y-S site leading to y_8 and a_2 ions being particularly

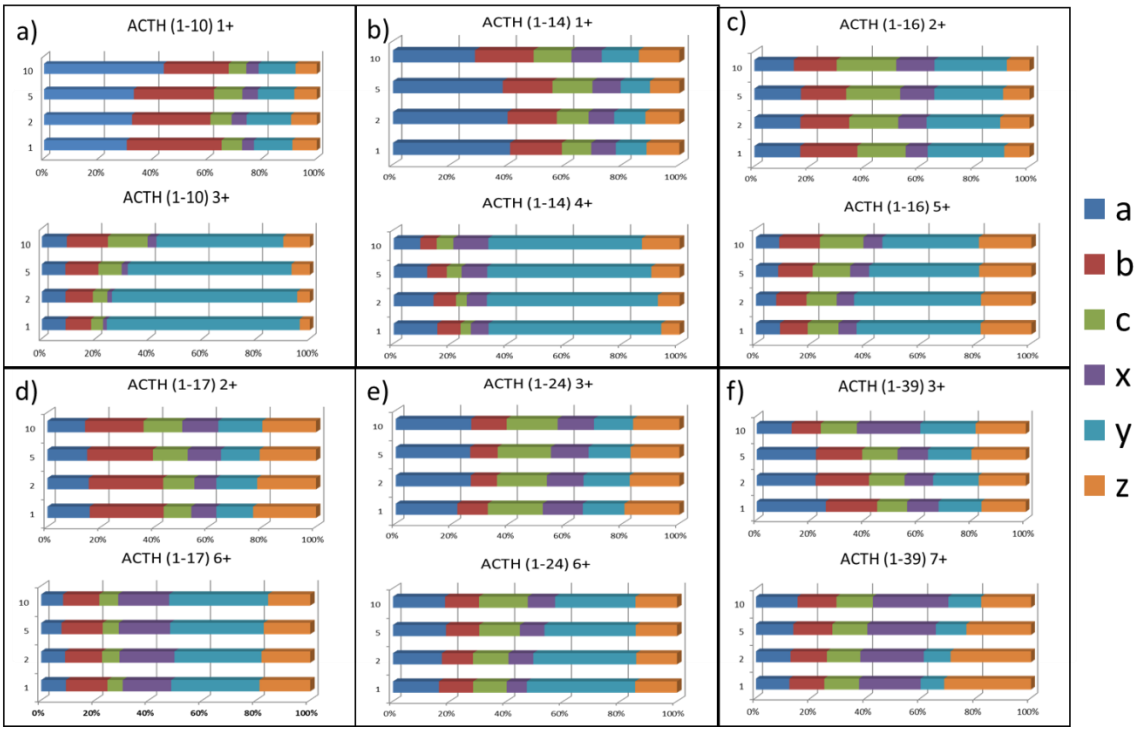


Figure 6.14: Distribution of 266 nm UVPD fragment ions for lowest and highest observed charge states of a) ACTH (1-10), b) ACTH (1-14), c) ACTH (1-16), d) ACTH (1-17), e) ACTH (1-24), and f) ACTH (1-39). The highest and lowest charge states were selected to show the shift to C-terminal ions as the charge state of the precursor increases. UVPD was carried out using 5 pulses at 6 mJ per pulse. The abundances of a, b, c, x, y, and z fragment ions were summed collectively based on ion type as a function of the number of laser pulses.

enhanced. The UVPD spectra exhibit special enhancement of *a* ions that are not favored in the CID spectra, and these may arise from secondary dissociation of *b* ions formed with excess internal energy or ones produced directly from the precursor peptide,

potentially via dissociation of ions in excited electronic states. The similarities between the 266 nm UVPD and 193 nm UVPD spectra (**Figures 6.1, 6.9** and **6.10**), both in terms of the identities and distributions of fragment ions, is remarkable. The level of similarities even extends to the formation of a few low abundance *z*-type fragment ions which are absent from the corresponding CID mass spectra. Apparently absorption of 266 nm (4.7 eV) or 193 nm (6.4 eV) photons allows access to excited states (whether the same ones or different ones for each of these wavelengths), and the activation process leads to similar fragmentation processes. Absorption of 193 nm photons may occur at the amide backbone as well as the aromatic side-chains of W, Y and F (the sites where 266 nm photons are absorbed), with W having a greater photoabsorption cross section than Y or F.[48] Based on previous studies of UVPD, it is believed that dissociation may occur directly from excited electronic states as well as after internal conversion and intramolecular vibrational redistribution.[39,68]Based on the present results that show a mixture of CID-like fragment ions (b/y) and radical-directed fragment ions (a/x, c/z), it appears that UVPD involves more than one type of process.

The UVPD and CID spectra collected for ACTH (1-14) showed an even more dramatic preferential cleavage of the Y-S bond, leading to the dominant y_{12}^{2+} ion for the 3+ and 4+ charge states (and complementary a_2 and b_2 ions). The lower charge states (1+, 2+) instead exhibited enhanced b_{11} and b_{13} ions upon 266 nm UVPD and CID, with the former representing a proline-type cleavage that also accounts for the prominent complementary y_3 ion. Interestingly, despite the overlap in the sequences of ACTH (1-10) and ACTH (1-14), the resulting MS/MS spectra had few similarities, an outcome that

is largely attributed to the C-terminal ion dominance in the spectra that masks the more subtle variations in the N-terminal ion abundances.

For ACTH (1-16), the Y-S cleavage (formation of y_{14}) remained the most consistently prominent process upon UVPD and CID for the higher charge states (3+, 4+, and 5+) (see **Figure 6.5**). Cleavage of the backbone between K-K (resulting in b_{15}) was also a significant pathway, which is a hallmark charge-modulated process common for both UVPD and CID. Selected *a, c,* and *z* ions were observed in the UVPD (266 nm and 193 nm) spectra which again reflected the greater diversity of pathways upon UV photoactivation.

ACTH (1-17) is an interesting example, having just a single additional arginine residue at the C-terminus compared to ACTH (1-16). The resulting UVPD mass spectra (**Figure 6.1**) exhibited many of the same preferential cleavage sites as observed for ACTH (1-16), such as the enhanced Y-S cleavage and cleavage adjacent to a basic residue. Similar to ACTH (1-16), the ACTH (1-17) ions in lower charge states (2+ and 3+) underwent a dominant cleavage of the last amino acid in the sequence (K-R bond) resulting in the formation the b_{16} ion. UVPD of ACTH (1-17) also show prominent formation of the y_{15} ions which corresponded to the Y-S cleavage for 3+, 4+, 5+, and 6+ charge states, as well as a modest increase in C-terminal ions as the charge state increased.

Four charge states (3+, 4+, 5+, and 6+) were observed for ACTH (1-24), and representative fragmentation patterns generated by UVPD and CID are shown in **Figure 6.6**. The relative portion of N-terminal ions compared to C-terminal ions produced by

UVPD was greater for all charge states compared to the distributions observed for the shorter ACTH peptides. Again the Y-S bond cleavage was particularly favored, yielding the y_{22} (and a_2) ions noted previously. The a_{22} ion, arising from a V-Y backbone cleavage, was a significant product ion upon UVPD, and it has been observed previously that cleavages C-terminal to valine and N-terminal to tyrosine may be enhanced.[66]

As the largest peptide in the series, ACTH (1-39) generated five charge states (3+, 4+, 5+,6+ and 7+) upon ESI and some rather unique results for both 266 nm UVPD and CID (**Figure 6.7**). This peptide sequence is unusual in that there are no basic sites found within the final 18 residues (C-terminus). For 266 nm UVPD the cleavage of the S-Y bond was not observed for any charge state; in fact there was no consistent preferential cleavage for any of the charge states upon UVPD. CID resulted in the dominant formation of a b_{35} ion which arises from cleavage of the F-P bond, a process consistent with the well-known rule of peptide cleavage N-terminal to proline. N-terminal *b* ions were highly favored upon CID of all charge states, presumably because the most basic charge sites which were clustered closer to the C-terminus for the shorter peptides, are actually situated in the mid-region of ACTH (1-39). This means that cleavages at backbone sites in the mid-section of the peptide more likely result in fragment ions with protons localized in the first 21 residues, thus enhancing N-terminal ions (such as the dominant b_{29}, b_{33}, b_{35}, b_{38} ions, among others, noted in **Table 6.1**). Interestingly, the distribution of N-terminal versus C-terminal ions for ACTH (1-39) is almost evenly split (50%/50%) upon 266 nm UVPD and 193 nm UVPD. In contrast, N-terminal fragments make up 70-75% of the fragment ions upon CID. The C-terminal ions that are observed

are long ones that contain basic residues from the N-terminal half of the peptide. The notable differences in C-terminal versus N-terminal fragment ion distribution further supports the lower degree of charge-site dependence on the fragmentation pathways observed upon UVPD.

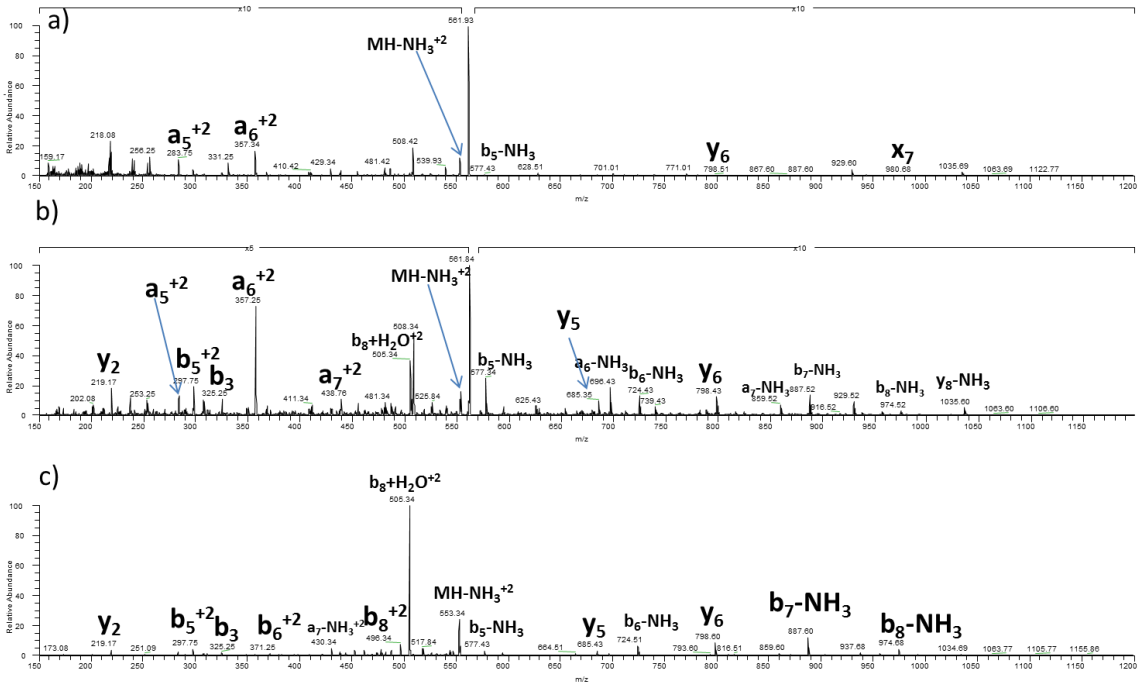


Figure 6.15: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of Alpha-Bag Cell Peptide. Precursor ion is in the 2+ charge state. a) 266 nm UVPD (5 pulses; 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

In an effort to elucidate the reason for the apparent preferential cleavage of the Y-S bond observed upon activation of the ACTH peptides, four additional peptides were subjected to CID, 193 UVPD, and 266 nm UVPD. The dominant Y-S cleavage could be

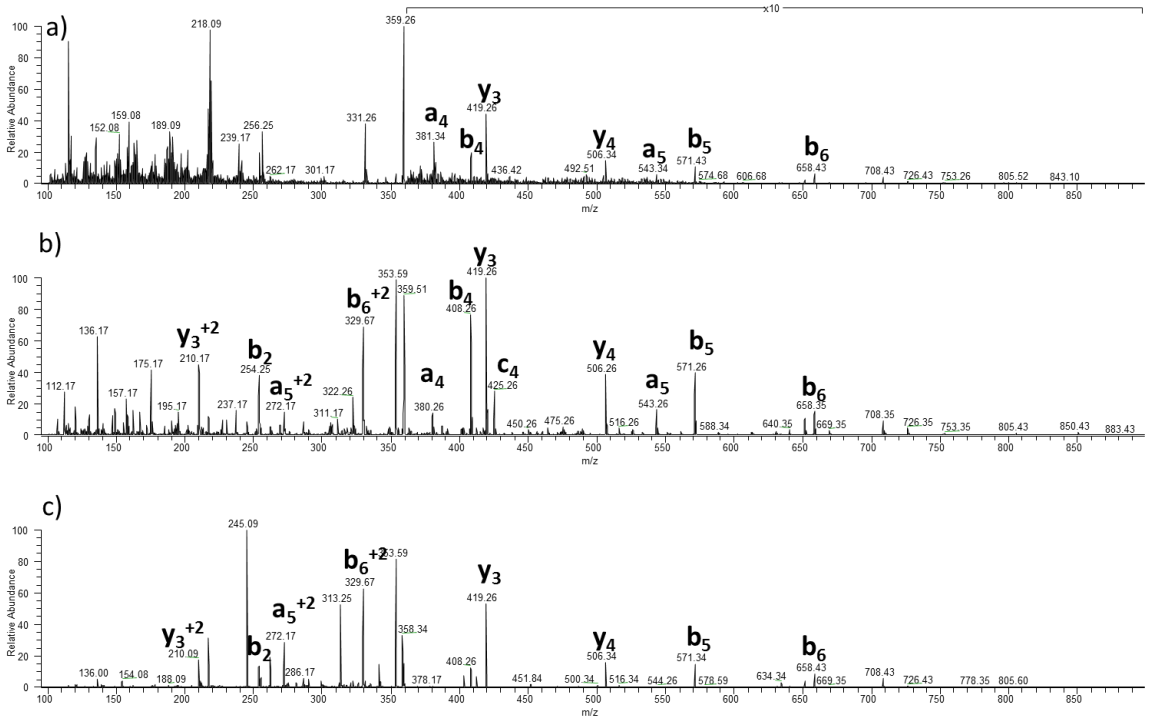


Figure 6.16: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of Bradykinin [Tyr 5]. Precursor ion is in the 3+ charge state. a) 266 nm UVPD (5 pulses; 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

due to its position in the peptide (the second and third residues), a special lability of the Y-S bond, or a chromophore effect (although it is noted that this bond cleavage is also observed upon CID, not just UVPD). Bradykinin [Tyr5] and alpha- bag cell peptide (1-9) were chosen because they both have a Y-S bond near the C-terminus (alpha-bag cell peptide (1-9) or the middle (bradykinin [Tyr5] of the peptide rather than near the N-terminus. For the alpha-bag cell peptide (1-9) fragments resulting from a cleavage of the Y-S bond appear as some of the more abundant fragments observed (**Figure 6.15**). The cleavage of the Y-S bond was fairly prominent for bradykinin[Tyr 5] and was only surpassed by proline cleavages (**Figure 6.16**), giving evidence that the preferential Y-S

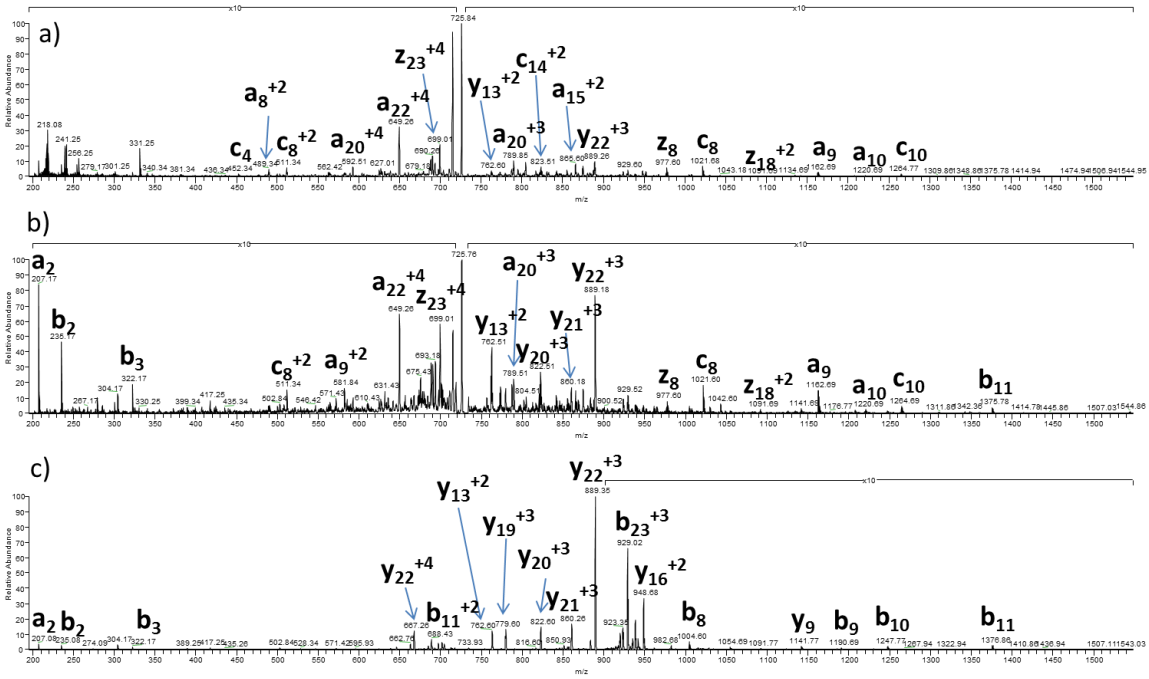


Figure 6.17: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of ACTH (1-24) [Phe 2, Nle 4]. Precursor ion is in the 4+ charge state. a) 266 nm UVPD (5 pulses; 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

cleavage observed for the ACTH peptide is not solely due to location of the Y-S residues near the N-terminus . Another peptide selected was ACTH (1-24) [Phe 2 Nle 4] for which the Tyr in ACTH at position 2 is replaced by Phe. The fragments observed from both UVPD and CID for this peptide are nearly identical to those seen for the original version of the ACTH (1-24) peptide, with the y_{22} and a_2 ions produced in great abundance (**Figure 6.17**). The fact that cleavage between the F-S residues parallels the preferential cleavage observed for the Y-S residues suggests that the preferential cleavage may arise from an aromatic residue/Ser motif. The third peptide analyzed was adamtsostatin-16

which contained a W and S residues at the third and fourth positions, respectively. The resulting UVPD and CID mass spectra are shown in **Figure 6.18**. In this case, cleavage of the W-S bond is not overly favored. In total, it appears that a new preferential cleavage associated with X-S in which X is Y or F occurs upon UVPD or CID, a feature that might be prevalent in larger scale bottom-up proteomic studies.

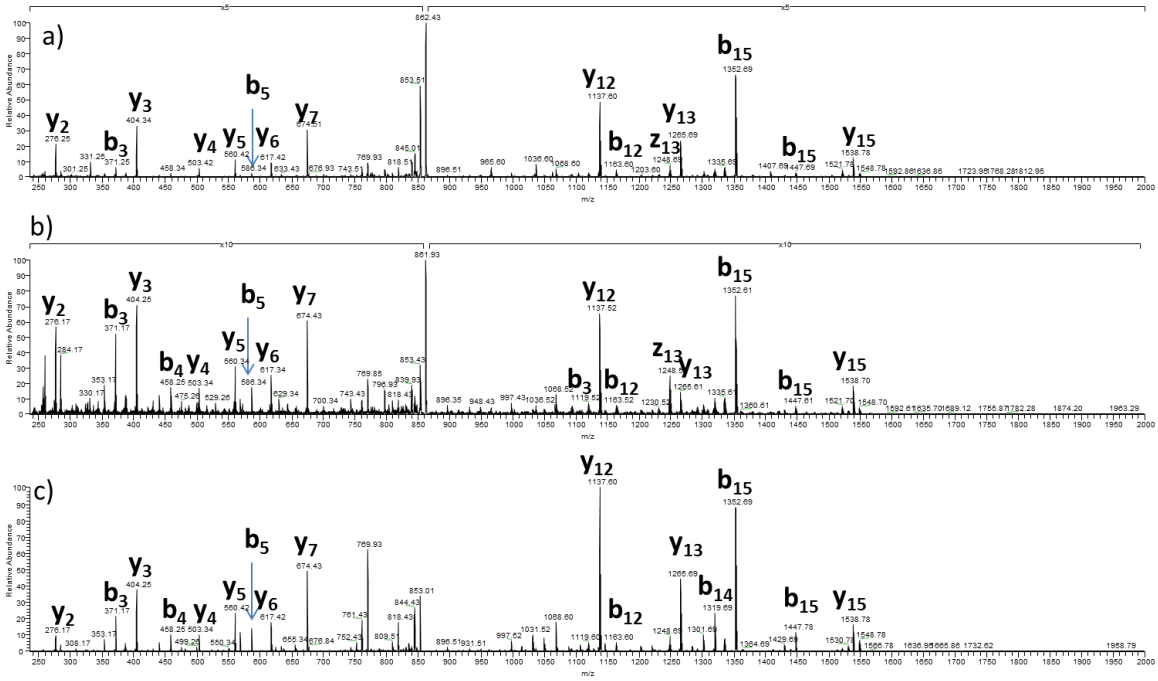


Figure 6.18: Examples of 266 nm UVPD, 193 nm UVPD, and CID mass spectra of Adamtsostatin-16. Precursor ion is in the 2+ charge state. a) 266 nm UVPD (5 pulses; 6 mJ per pulse), b) 193 nm UVPD (2 pulses; 2 mJ per pulse), and c) CID (NCE 35%).

6.4.3 Comparison of 266 nm, 193 nm, CID sequence coverage for ACTH peptides

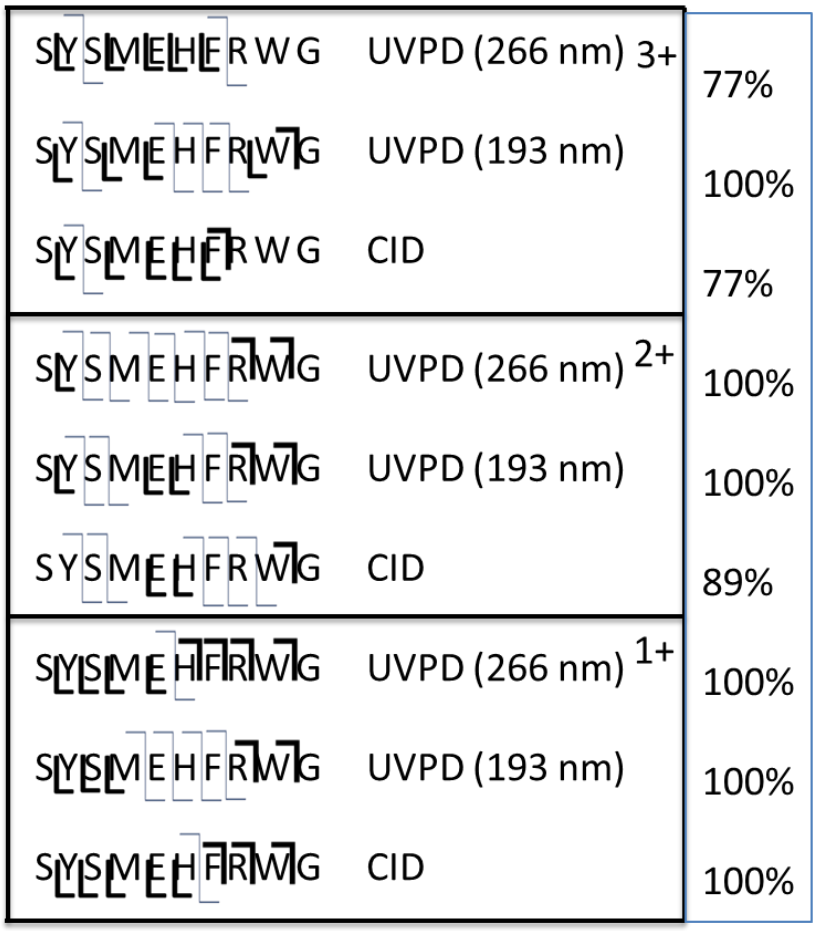


Figure 6.19: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-10). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by ⌊ ⌋.

Among the series of ACTH peptides, the UVPD spectra of the smaller peptides were more "CID-like" in terms of the portion of *a,b*, and *y* ions, particularly for ACTH (1-10) (3+) and ACTH (1-14) (4+). For the longer ACTH peptides and/or ones in lower

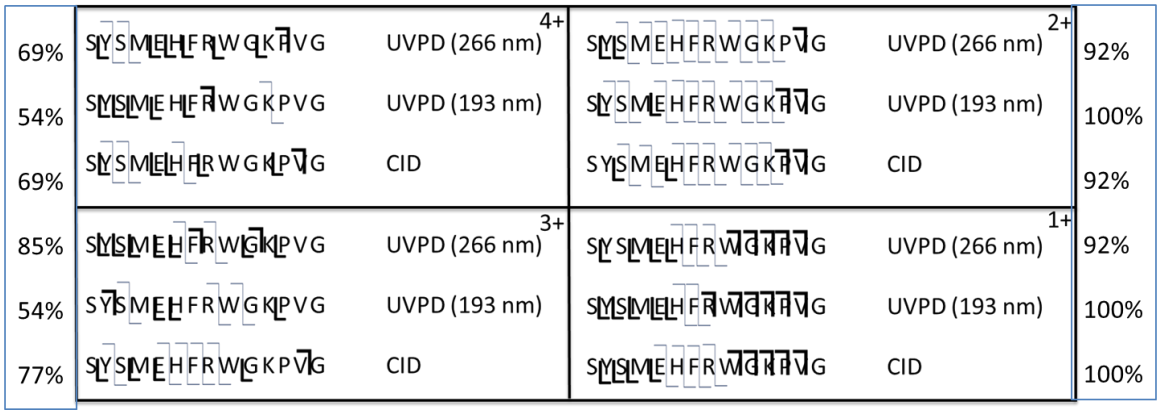


Figure 6.20: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-14). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by

charge states, the UVPD spectra exhibited the greater diversity of fragment ions that is the established hallmark of UV photoactivation. In terms of sequence coverages (i.e. expressed as a percentage based on the number of backbone cleavages relative to the total number of backbone positions), UVPD and CID typically gave similar coverages for the lower charge states of each peptide (see values in **Figures 6.18 – 23**).

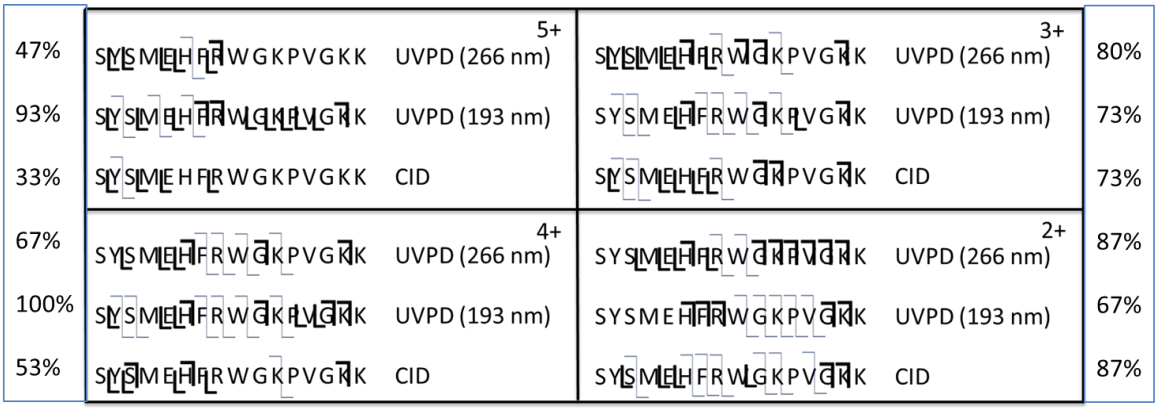


Figure 6.21: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-16). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by ┌ ┘.

However, for higher charge states, UVPD generally yielded higher coverages compared to CID, especially for 193 nm UVPD. For example, the 2+ charge state of ACTH (1-16) yielded a sequence coverage of 87% (13 backbone cleavage sites out of 15 sites) for both CID and 266 nm UVPD. However, for the 5+ charge state the sequence coverage decreased to only 33% for CID and 47% for 266 nm UVPD. The sequence coverage obtained upon 193 nm UVPD showed less variation, ranging from 67% to 100% across the four charge states. For ACTH (1-17), **Figure 6.21** shows a sequence coverage map for all five charges states for 193 nm, 266 nm, and CID fragmentation.

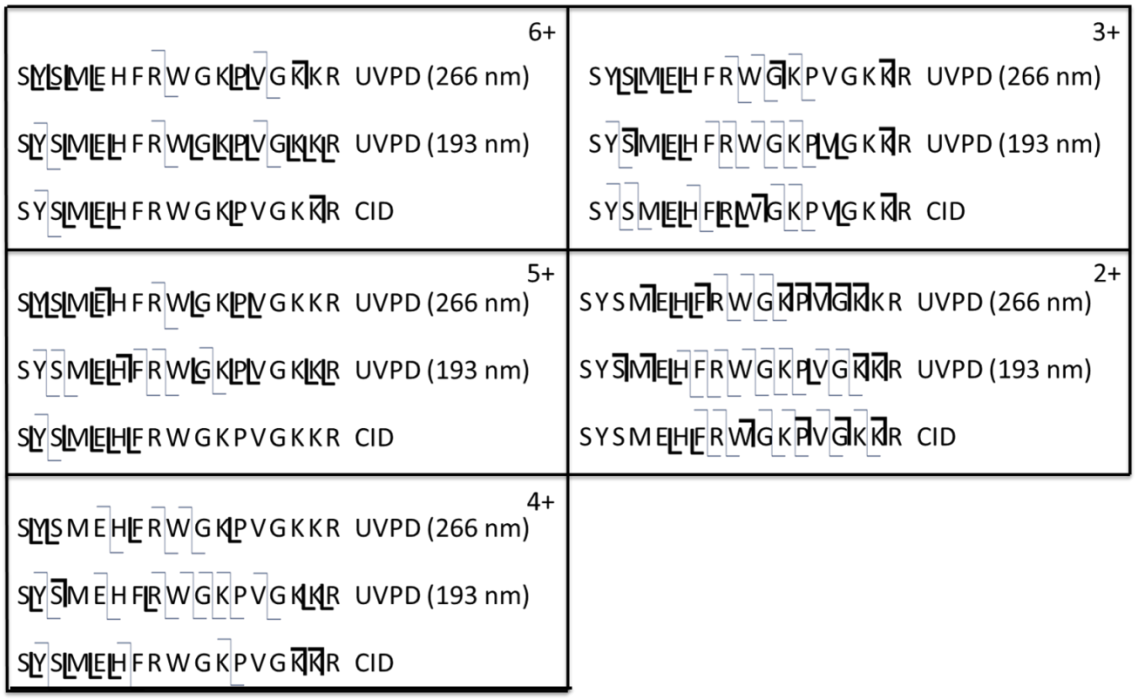


Figure 6.22: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-16). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by ┌ ┘.

A maximum of 75% coverage was obtained by CID and 266 nm UVPD for the lowest 2+ charge state, whereas the coverage decreased to 38% for CID and 56% for 266 nm UVPD of the highest 6+ charge state. The sequence coverage produced by 193 nm UVPD was again less charge-state dependent, ranging from 75% to 88% for ACTH (1-17). As anticipated, the level of coverage varied more dramatically for CID as a function of charge state compared to UVPD. For instance, the sequence coverage ranged from 70% (3+) to 26% (6+) for CID of the ACTH (1-24) peptide, whereas the range varied from 65% (3+) to 52% (6+) for 266 nm UVPD and from 74% (3+) to 61% (6+) for 193

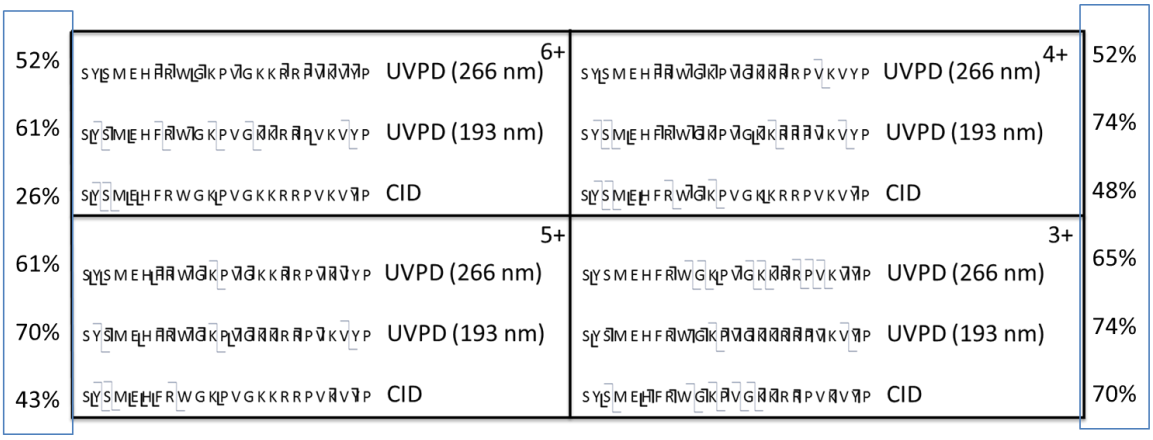


Figure 6.23: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-24). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by ⌊ ⌋.

nm UVPD of the same peptide. The variation in sequence coverage was most notable for the longer ACTH peptides, those that contained basic lysine or arginine residues at the C-terminus or in the middle of the sequences. The change in sequence coverage obtained by CID for the longer peptides as a function of charge state was not surprising given the fact that CID is more highly modulated by proton mobility than UVPD.

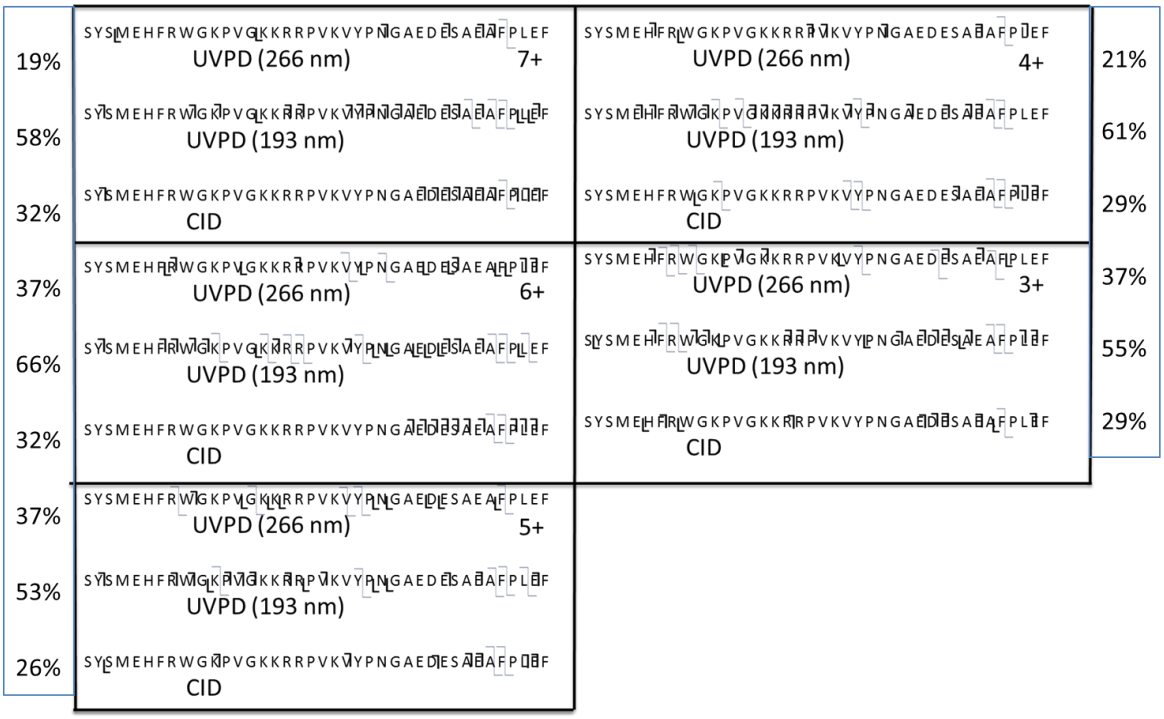


Figure 6.24: Sequence coverage maps for observed ions in all observed charge states of ACTH (1-39). Any backbone cleavage site producing N-terminal a, b, c ions is represented by ┐. Any backbone cleavage producing C-terminal x,y, z ions is represented by └. Cleavages that result in both N-terminal a, b, c ions and C-terminal x, y, z ions are designated by ┌ ┘.

6.5 CONCLUSIONS

266 nm UVPD and 193 nm UVPD generated similar fragment ion distributions for each ACTH peptide, spanning a variety of *a,b,c* and *x,y,z* ions. Comparison of the MS/MS results for ACTH (1-10), ACTH (1-14), ACTH (1-16), ACTH (1-17), and ACTH (1-24) revealed that UVPD and CID consistently showed preferential cleavage of the Y-S bond for nearly every charge state. For all of the peptides, the production of C-terminal versus N-terminal ions and overall sequence coverage was far more dependent on the charge state and location of basic sites for CID than for UVPD, an outcome that reflects

the reduced prominence of charge-mediated pathways for UVPD. In general, UVPD of the longer peptides (ACTH (1-16), ACTH (1-17), ACTH (1-24) and ACTH (1-39)) showed relatively little change in the distributions of fragment ion types as a function of the charge state of the peptide; there were much greater changes observed for CID. UVPD demonstrated a modest degree of preferential cleavage adjacent to amino acids containing aromatic side-chains, suggestive of a chromophore effect, as well as enhanced cleavages adjacent to proline akin to the well-known proline effect documented for CID. This systematic comparative study has demonstrated many similarities in the types and distributions of fragment ions produced by 266 nm UVPD and 193 nm UVPD, with greater overall sequence coverage afforded by 193 nm UVPD. A comparison of the results for 193 nm and 266 nm UVPD shows that shows that both provide better sequence coverage compared to CID for higher charge states of larger peptides. 193 nm UVPD provides better coverage on average than 266 nm UVPD, and the sequence coverages obtained by UVPD did not exhibit the significant dependence on charge state that was observed upon CID.

6.6 REFERENCES

(1) Swaney, D. L.; McAlister, G. C.; Coon, J. J. Decision Tree–driven Tandem Mass Spectrometry for Shotgun Proteomics. *Nat. Methods* **2008**, *5* (11), 959–964.

(2) Bailey, D. J.; Rose, C. M.; McAlister, G. C.; Brumbaugh, J.; Yu, P.; Wenger, C. D.; Westphall, M. S.; Thomson, J. A.; Coon, J. J. Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2012**, *109* (22), 8411–8416.

(3) Pan, C.; Zhou, Y.; Dator, R.; Ginghina, C.; Zhao, Y.; Movius, J.; Peskind, E.; Zabetian, C. P.; Quinn, J.; Galasko, D.; Stewart, T.; Shi, M.; Zhang, J. Targeted Discovery and Validation of Plasma Biomarkers of Parkinson's Disease. *J. Proteome Res.* **2014**.

(4) Ma, J.; Ward, C. C.; Jungreis, I.; Slavoff, S. A.; Schwaid, A. G.; Neveu, J.; Budnik, B. A.; Kellis, M.; Saghatelian, A. Discovery of Human sORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**, *13* (3), 1757–1765.

(5) Seneviratne, U.; Godoy, L. C.; Wishnok, J. S.; Wogan, G. N.; Tannenbaum, S. R. Mechanism-Based Triarylphosphine-Ester Probes for Capture of Endogenous RSNOs. *J. Am. Chem. Soc.* **2013**, *135* (20), 7693–7704.

(6) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(7) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F. M.; Mechtler, K.; Heck, A. J. R.; Mohammed, S. Unambiguous Phosphosite Localization Using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *J. Proteome Res.* **2013**, *12* (3), 1520–1525.

(8) Mommen, G. P. M.; Frese, C. K.; Meiring, H. D.; Brink, J. van G. den; Jong, A. P. J. M. de; Els, C. A. C. M. van; Heck, A. J. R. Expanding the Detectable HLA Peptide Repertoire Using Electron-Transfer/higher-Energy Collision Dissociation (EThcD). *Proc. Natl. Acad. Sci.* **2014**, *111* (12), 4507–4512.

(9) Gunawardena, H. P.; He, M.; Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; Hodges, B. D. M.; McLuckey, S. A. Electron Transfer versus Proton Transfer in Gas-Phase Ion/Ion Reactions of Polyprotonated Peptides. *J. Am. Chem. Soc.* **2005**, *127* (36), 12627–12639.

(10) McGee, W. M.; McLuckey, S. A. The Ornithine Effect in Peptide Cation Dissociation: The Ornithine Effect: Selective Cleavages Observed in Ornithinated Peptides. *J. Mass Spectrom.* **2013**, *48* (7), 856–861.

(11) Krusemark, C. J.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. Modifying the Charge State Distribution of Proteins in Electrospray Ionization Mass Spectrometry by Chemical Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (9), 1617–1625.

(12) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (24), 9243–9252.

(13) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(14) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(15) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. Mass Spectrometry of Peptides and Proteins. *Methods* **2005**, *35* (3), 211–222.

(16) Beys-da-Silva, W. O.; Santi, L.; Berger, M.; Calzolari, D.; Passos, D. O.; Guimarães, J. A.; Moresco, J. J.; Yates, J. R. Secretome of the Biocontrol Agent Metarhizium Anisopliae Induced by the Cuticle of the Cotton Pest Dysdercus Peruvianus Reveals New Insights into Infection. *J. Proteome Res.* **2014**, *13* (5), 2282–2296.

(17) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(18) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(19) Peng, Y.; Chen, X.; Zhang, H.; Xu, Q.; Hacker, T. A.; Ge, Y. Top-down Targeted Proteomics for Deep Sequencing of Tropomyosin Isoforms. *J. Proteome Res.* **2013**, *12* (1), 187–198.

(20) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(21) Rose, C. M.; Russell, J. D.; Ledvina, A. R.; McAlister, G. C.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J.; Syka, J. E. P. Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (6), 816–827.

(22) Laskin, J.; Futrell, J. H. Surface-Induced Dissociation of Peptide Ions: Kinetics and Dynamics. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (12), 1340–1347.

(23) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 190–208.

(24) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(25) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(26) Chingin, K.; Makarov, A.; Denisov, E.; Rebrov, O.; Zubarev, R. A. Fragmentation of Positively-Charged Biological Ions Activated with a Beam of High-Energy Cations. *Anal. Chem.* **2014**, *86* (1), 372–379.

(27) Cook, S. L.; Collin, O. L.; Jackson, G. P. Metastable Atom-Activated Dissociation Mass Spectrometry: Leucine/isoleucine Differentiation and Ring Cleavage of Proline Residues. *J. Mass Spectrom.* **2009**, *44* (8), 1211–1223.

(28) Flora, J. W.; Muddiman, D. C. Determination of the Relative Energies of Activation for the Dissociation of Aromatic versus Aliphatic Phosphopeptides by ESI-FTICR-MS and IRMPD. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (1), 121–127.

(29) Brodbelt, J. S.; Wilson, J. J. Infrared Multiphoton Dissociation in Quadrupole Ion Traps. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.

(30) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(31) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

(32) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(33) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(34) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(35) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(36) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.

(37) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(38) Zhang, L.; Reilly, J. P. Radical-Driven Dissociation of Odd-Electron Peptide Radical Ions Produced in 157 Nm Photodissociation. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (7), 1378–1390.

(39) Parthasarathi, R.; He, Y.; Reilly, J. P.; Raghavachari, K. New Insights into the Vacuum UV Photodissociation of Peptides. *J. Am. Chem. Soc.* **2010**, *132* (5), 1606–1610.

(40) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide Ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.

(41) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 Nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19* (12), 1657–1665.

(42) Wilson, J. J.; Brodbelt, J. S. Ultraviolet Photodissociation at 355 Nm of Fluorescently Labeled Oligosaccharides. *Anal. Chem.* **2008**, *80* (13), 5186–5196.

(43) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(44) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(45) Aponte, J. R.; Vasicek, L.; Swaminathan, J.; Xu, H.; Koag, M. C.; Lee, S.; Brodbelt, J. S. Streamlining Bottom-up Protein Identification Based on Selective Ultraviolet Photodissociation (UVPD) of Chromophore-Tagged Histidine- and Tyrosine-Containing Peptides. *Anal. Chem.* **2014**, *86* (13), 6237–6244.

(46) Joly, L.; Antoine, R.; Broyer, M.; Dugourd, P.; Lemoine, J. Specific UV Photodissociation of Tyrosyl-Containing Peptides in Multistage Mass Spectrometry. *J. Mass Spectrom.* **2007**, *42* (6), 818–824.

(47) Oh, J. Y.; Moon, J. H.; Kim, M. S. Sequence- and Site-Specific Photodissociation at 266 Nm of Protonated Synthetic Polypeptides Containing a Tryptophanyl Residue. *Rapid Commun. Mass Spectrom.* **2004**, *18* (22), 2706–2712.

(48) Oh, J. Y.; Moon, J. H.; Kim, M. S. Chromophore Effect in Photodissociation at 266 Nm of Protonated Peptides Generated by Matrix-Assisted Laser Desorption Ionization (MALDI). *J. Mass Spectrom.* **2005**, *40* (7), 899–907.

(49) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. Photodissociation Tandem Mass Spectrometry at 266 Nm of an Aliphatic Peptide Derivatized with Phenyl Isothiocyanate and 4-Sulfophenyl Isothiocyanate. *Rapid Commun. Mass Spectrom.* **2005**, *19* (10), 1283–1288.

(50) Park, S.; Ahn, W.-K.; Lee, S.; Han, S. Y.; Rhee, B. K.; Oh, H. B. Ultraviolet Photodissociation at 266 Nm of Phosphorylated Peptide Cations. *Rapid Commun. Mass Spectrom.* **2009**, *23* (23), 3609–3620.

(51) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(52) Diedrich, J. K.; Julian, R. R. Site Selective Fragmentation of Peptides and Proteins at Quinone Modified Cysteine Residues Investigated by ESI-MS. *Anal. Chem.* **2010**, *82* (10), 4006–4014.

(53) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358.

(54) Liu, Z.; Julian, R. R. Deciphering the Peptide Iodination Code: Influence on Subsequent Gas-Phase Radical Generation with Photodissociation ESI-MS. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (6), 965–971.

(55) Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609.

(56) Diedrich, J. K.; Julian, R. R. Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. *Anal. Chem.* **2011**, *83* (17), 6818–6826.

(57) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of Phosphorylation Site-Specific Dissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.

(58) Sun, Q.; Yin, S.; Loo, J. A.; Julian, R. R. Radical Directed Dissociation for Facile Identification of Iodotyrosine Residues Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2010**, *82* (9), 3826–3833.

(59) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(60) Enjalbert, Q.; Simon, R.; Salvador, A.; Antoine, R.; Redon, S.; Ayhan, M. M.; Darbour, F.; Chambert, S.; Bretonnière, Y.; Dugourd, P.; Lemoine, J. Photo-SRM: Laser-Induced Dissociation Improves Detection Selectivity of Selected Reaction Monitoring Mode. *Rapid Commun. Mass Spectrom.* **2011**, *25* (22), 3375–3381.

(61) Enjalbert, Q.; Girod, M.; Simon, R.; Jeudy, J.; Chirot, F.; Salvador, A.; Antoine, R.; Dugourd, P.; Lemoine, J. Improved Detection Specificity for Plasma Proteins by Targeting Cysteine-Containing Peptides with Photo-SRM. *Anal. Bioanal. Chem.* **2013**, *405* (7), 2321–2331.

(62) Girod, M.; Biarc, J.; Enjalbert, Q.; Salvador, A.; Antoine, R.; Dugourd, P.; Lemoine, J. Implementing Visible 473 Nm Photodissociation in a Q-Exactive Mass Spectrometer: Towards Specific Detection of Cysteine-Containing Peptides. *Analyst* **2014**, *139* (21), 5523–5530.

(63) Tecklenburg, R. E.; Miller, M. N.; Russell, D. H. Laser Ion Beam Photodissociation Studies of Model Amino Acids and Peptides. *J. Am. Chem. Soc.* **1989**, *111* (4), 1161–1171.

(64) Solouki, T.; Russell, D. H. Structural Mass Spectrometry of Matrix-Assisted Laser-Desorbed Biomolecules by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Photoionization and Photofragmentation. *Appl. Spectrosc.* **1993**, *47* (2), 211–217.

(65) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(66) Lai, C.-K.; Ng, D. C. M.; Pang, H. F.; Le Blanc, J. C. Y.; Hager, J. W.; Fang, D.-C.; Cheung, A. S.-C.; Chu, I. K. Laser-Induced Dissociation of Singly Protonated Peptides at 193 and 266 Nm within a Hybrid Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2013**, *27* (10), 1119–1127.

(67) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom. JMS* **2000**, *35* (12), 1399–1406.

(68) Hu, Y.; Hadas, B.; Davidovitz, M.; Balta, B.; Lifshitz, C. Does IVR Take Place Prior to Peptide Ion Dissociation? *J. Phys. Chem. A* **2003**, *107* (34), 6507–6514.

Chapter 7

Conclusions

The use of database search tools to evaluate mass spectrometry results has contributed to tremendous advancements in the understanding of the complex field of proteomics. In spite of all of the discoveries there is still a great deal about proteomics that remains unknown as it is almost impossible to predict the exact modifications or mutations that proteins may undergo. In an effort to better characterize and pinpoint the alterations protein experience, *de novo* sequencing programs are implemented which allow elucidation of protein information without any prior knowledge of the sample. The focus of this dissertation is the development of novel *de novo* methods which use peptide derivatization as well as ultraviolet photodissociation (UVPD) to create simplified spectra and subsequently employ those spectra to use a *de novo* program which capitalizes on the simplification of these spectra to yield superior *de nov*o results.

In Chapter 3, tryptic peptides were derivatized with sulfosuccinimydyl-7-amino-4-methyl-coumarin-3-acetic acid (AMCA), a UV chromophore, and then activated using 351 nm UVPD. Fragmentation with 351 nm UVPD caused the elimination of *b*-ions so that only y-ions remained as the lone series of ions observed. The simplified spectra were shown to be perfectly suited for *de novo* sequencing. The implementation of technique was shown to be effective in the analysis of single proteins, protein mixtures, as well as the differentiation of several different green fluorescent protein (GFP) variants.

The improvement of the AMCA/351 nm method was discussed in Chapter 4. Refinements to the experimental workflow allowed this *de novo* technique to be applied

to complex cell lysates, creating simplified spectra from the AMCA/351 nm UVPD strategy. Using these spectra from the cell lysate a novel *de novo* sequencing program created by Andrew Horton that processed spectra using a Random Forest model and then generated potential peptide sequences for the spectra using a hidden Markov model (HMM) was tested. This new method was able to identify thousands of peptides from an *E.coli* lysate with high confidence and yielded results comparable or superior to that of other *de novo* algorithms.

Further modifications were made to the new *de novo* program in Chapter 5. Paired UVPD/CID datasets in conjunction with the *de novo* algorithm resulted in an improvement in peptide identification. This improvement was due in large part to the fact that the UVnovo program exploited the individual attributes of UVPD and CID spectra in a way that mitigated the shortcomings of either CID or UVPD spectra alone. In addition to incorporation of paired spectra, the *de novo* program was also modified to allow the use of sequence tags that were matched to a known database so the program could be used as a hybrid *de novo*/database search method.

In Chapter 6, peptide characterization of six peptides wass performed using 266 nm UVPD. Fragmentation data collected for these peptides was compared to collision induced dissociation (CID) and 193 nm UVPD to better understand characteristics and pathways of 266 nm UVPD fragmentation based on charge state and amino acid composition.

In summary, UVPD was utilized for peptide characterization or used in tandem with derivatization of peptides to attach UV chromophores to generate spectra ideal for

de novo sequencing by eliminating b-ions from MS/MS spectra. A *de novo* sequencing algorithm was used to maximize the utility of the unique fragmentation patterns produced by UVPD, allowing thousands of peptides to be confidently sequenced. The novel de novo method also allowed paired UVPD/CID spectra to be searched which allowed it to be used to match peptides to a known database by the implementation of sequence tags.

Future effort could be envisioned in two areas. First, there would be huge benefits to using the novel 351 nm UVPD *de novo* method for high accuracy mass spectra. The use of high resolution/high accuracy spectra will afford significant improvement in the assignment of the precursor mass of peptides which plays a crucial role in the ability of the *de novo* program to assign the correct composition and number of amino acids in a peptide. Furthermore, high mass accuracy greatly reduces the possible number of sequence combinations feasible for a precursor mass, thus reducing the ambiguity in sequence assignment. The use of high resolution data will also allow differentiation of amino acid pairs from single amino acids which share the same nominal mass. For example a loss of a GV pair of amino acids from a peptide sequence has a similar nominal mass as a loss of R based on low resolution mass analysis, whereas high resolution data can easily distinguish between the two. With all of these benefits the use of high resolution in tandem with this 351 nm *de novo* program should result in a significant improvement in the confidence of *de novo* sequencing results.

Secondly, the use of the AMCA/351 nm UVPD method being applied to data independent acquisition (DIA) strategies is a compelling target of opportunity. DIA is a method for data collection where all peptides in a defined *m/z* range (such as a 25 Da

interval) are activated and dissociated simultaneously (rather than one selected ion at a time), allowing several peptides to be identified at once. Although the resulting MS/MS pattern is a conglomeration of all the fragment ions from all the peptide precursors in the 25 Da window, identification of these peptides is typically performed by an extensive computational reconstruction of fragmentation spectra from complex data sets. The use of DIA has gained significant interest in recent years as it does not suffer from a bias in peak selection process, a common sampling problem in data dependent acquisition (DDA) in which individual peptide masses are isolated and activated to generate MS/MS spectra. For DDA methods, typically the ten most abundant ions observed in an MS1 spectra are sequentially selected for isolation and activation. The bias in peak selections results in low abundance ions often being missed as the program overlooks them in favor of more abundant peaks for analysis. Due to the fact that DIA simultaneously activates and dissociates multiple peptides within a *m/z* range, it currently faces the problem of processing extremely complex data from congested, conglomerated MS/MS spectra. I believe the AMCA/351 nm UVPD method could overcome some of the problems associated with DIA strategies. The fragmentation of peptides with the 351 nm UVPD method greatly reduces the complexity of MS/MS spectra being analyzed. The analysis of these simplified spectra with a DIA program method should streamline the spectral deconvolution process and lead to more confident and accurate peptide assignments.

REFERENCES

Chapter 1

(1) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312* (5771), 212–217.

(2) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(4) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(5) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(6) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(7) Xu, H.; Freitas, M. A Mass Accuracy Sensitive Probability Based Scoring Algorithm for Database Searching of Tandem Mass Spectrometry Data. *BMC Bioinformatics* **2007**, *8* (1), 133.

(8) Craig, R.; Beavis, R. C. A Method for Reducing the Time Required to Match Protein Sequences with Tandem Mass Spectra. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2310–2316.

(9) Balgley, B. M.; Laudeman, T.; Yang, L.; Song, T.; Lee, C. S. Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. *Mol. Cell. Proteomics* **2007**, *6* (9), 1599–1608.

(10) De Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. Comprehensive Mass-Spectrometry-Based Proteome Quantification of Haploid versus Diploid Yeast. *Nature* **2008**, *455* (7217), 1251–1254.

(11) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(12) Standing, K. G. Peptide and Protein de Novo Sequencing by Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 595–601.

(13) Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.

(14) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(15) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(16) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(17) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(18) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(19) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(20) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.

(21) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(22) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(23) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533.

(24) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(25) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198–207.

(26) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39* (10), 1091–1112.

(27) Steen, H.; Fernandez, M.; Ghaffari, S.; Pandey, A.; Mann, M. Phosphotyrosine Mapping in Bcr/Abl Oncoprotein Using Phosphotyrosine-Specific Immonium Ion Scanning. *Mol. Cell. Proteomics* **2003**, *2* (3), 138–145.

(28) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat. Methods* **2007**, *4* (9), 709–712.

(29) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 190–208.

(30) Fernández, F. M.; Smith, L. L.; Kuppannan, K.; Yang, X.; Wysocki, V. H. Peptide Sequencing Using a Patchwork Approach and Surface-Induced Dissociation in Sector-TOF and Dual Quadrupole Mass Spectrometers. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (12), 1387–1401.

(31) Dongré, A. R.; Somogyi, A.; Wysocki, V. H. Surface-Induced Dissociation: An Effective Tool to Probe Structure, Energetics and Fragmentation Mechanisms of Protonated Peptides. *J. Mass Spectrom. JMS* **1996**, *31* (4), 339–350.

(32) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.

(33) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. Unraveling Molecular Complexity of Phosphorylated Human Cardiac Troponin I by Top down Electron Capture Dissociation/electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics MCP* **2008**, *7* (10), 1838–1849.

(34) Sweet, S. M. M.; Mardakheh, F. K.; Ryan, K. J. P.; Langton, A. J.; Heath, J. K.; Cooper, H. J. Targeted Online Liquid Chromatography Electron Capture Dissociation Mass Spectrometry for the Localization of Sites of in Vivo Phosphorylation in Human Sprouty2. *Anal. Chem.* **2008**, *80* (17), 6650–6657.

(35) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (24), 9243–9252.

(36) Crowe, M. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation (IRMPD) and Collisionally Activated Dissociation of Peptides in a Quadrupole Ion Trap with Selective IRMPD of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (11), 1581–1592.

(37) Payne, A. H.; Glish, G. L. Thermally Assisted Infrared Multiphoton Photodissociation in a Quadrupole Ion Trap. *Anal. Chem.* **2001**, *73* (15), 3542–3548.

(38) Bowers, W. D.; Delbert, S. S.; Hunter, R. L.; McIver, R. T. Fragmentation of Oligopeptide Ions Using Ultraviolet Laser Radiation and Fourier Transform Mass Spectrometry. *J. Am. Chem. Soc.* **1984**, *106* (23), 7288–7289.

(39) Hunt, D. F.; Shabanowitz, J.; Yates, J. R. Peptide Sequence Analysis by Laser Photodissociation Fourier Transform Mass Spectrometry. *J. Chem. Soc. Chem. Commun.* **1987**, No. 8, 548–550.

(40) Lebrilla, C. B.; Wang, D. T. S.; Mizoguchi, T. J.; McIver, R. T. Comparison of the Fragmentation Produced by Fast Atom Bombardment and Photodissociation of Peptides. *J. Am. Chem. Soc.* **1989**, *111* (23), 8593–8598.

(41) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 Nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19* (12), 1657–1665.

(42) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide Ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.

(43) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(44) Kim, T.-Y.; Reilly, J. P. Time-Resolved Observation of Product Ions Generated by 157 Nm Photodissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (12), 2334–2341.

(45) Moon, J. H.; Shin, Y. S.; Kim, M. S. Utility of Reaction Intermediate Monitoring with Photodissociation Multi-Stage (MSn) Time-of-Flight Mass Spectrometry for Mechanistic and Structural Studies: Phosphopeptides. *Int. J. Mass Spectrom.* **2009**, *288* (1–3), 16–21.

(46) Moon, J. H.; Yoon, S. H.; Bae, Y. J.; Kim, M. S. Dissociation Kinetics of Singly Protonated Leucine Enkephalin Investigated by Time-Resolved Photodissociation Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (7), 1151–1158.

(47) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(48) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(49) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(50) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein

Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(51) O'Brien, J. P.; Mayberry, L. K.; Murphy, P. A.; Browning, K. S.; Brodbelt, J. S. Evaluating the Conformation and Binding Interface of Cap-Binding Proteins and Complexes via Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **2013**, *12* (12), 5867–5877.

(52) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.

(53) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (37), 12920–12928.

(54) Cammarata, M.; Lin, K.-Y.; Pruet, J.; Liu, H.-W.; Brodbelt, J. Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-down Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86* (5), 2534–2542.

(55) Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D. Evolving tRNASec for Efficient Canonical Incorporation of Selenocysteine. *J. Am. Chem. Soc.* **2014**, *137* (1), 46–49.

(56) Cannon, J. R.; Martinez-Fonts, K.; Robotham, S. A.; Matouschek, A.; Brodbelt, J. S. Top-Down 193-Nm Ultraviolet Photodissociation Mass Spectrometry for

Simultaneous Determination of Polyubiquitin Chain Length and Topology. *Anal. Chem.* **2015**.

(57) Oh, J. Y.; Moon, J. H.; Kim, M. S. Sequence- and Site-Specific Photodissociation at 266 Nm of Protonated Synthetic Polypeptides Containing a Tryptophanyl Residue. *Rapid Commun. Mass Spectrom.* **2004**, *18* (22), 2706–2712.

(58) Oh, J. Y.; Moon, J. H.; Kim, M. S. Chromophore Effect in Photodissociation at 266 Nm of Protonated Peptides Generated by Matrix-Assisted Laser Desorption Ionization (MALDI). *J. Mass Spectrom.* **2005**, *40* (7), 899–907.

(59) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. Photodissociation Tandem Mass Spectrometry at 266 Nm of an Aliphatic Peptide Derivatized with Phenyl Isothiocyanate and 4-Sulfophenyl Isothiocyanate. *Rapid Commun. Mass Spectrom.* **2005**, *19* (10), 1283–1288.

(60) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(61) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358.

(62) Liu, Z.; Julian, R. R. Deciphering the Peptide Iodination Code: Influence on Subsequent Gas-Phase Radical Generation with Photodissociation ESI-MS. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (6), 965–971.

(63) Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609.

(64) Sun, Q.; Yin, S.; Loo, J. A.; Julian, R. R. Radical Directed Dissociation for Facile Identification of Iodotyrosine Residues Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2010**, *82* (9), 3826–3833.

(65) Diedrich, J. K.; Julian, R. R. Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. *Anal. Chem.* **2011**, *83* (17), 6818–6826.

(66) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(67) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(68) O'Brien, J. P.; Pruet, J. M.; Brodbelt, J. S. Chromogenic Chemical Probe for Protein Structural Characterization via Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (15), 7391–7397.

(69) Aponte, J. R.; Vasicek, L.; Swaminathan, J.; Xu, H.; Koag, M. C.; Lee, S.; Brodbelt, J. S. Streamlining Bottom-up Protein Identification Based on Selective Ultraviolet Photodissociation (UVPD) of Chromophore-Tagged Histidine- and Tyrosine-Containing Peptides. *Anal. Chem.* **2014**, *86* (13), 6237–6244.

(70) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(71) Zaikin, V.; Halket, J. M. *A Handbook of Derivatives for Mass Spectrometry*; IM Publications, 2009.

(72) Roth, K. D. W.; Huang, Z.; Sadagopan, N.; Watson, J. T. Charge Derivatization of Peptides for Analysis by Mass Spectrometry. *Mass Spectrom. Rev.* **1998**, *17* (4), 255–274.

(73) Vasicek, L.; Brodbelt, J. S. Enhanced Electron Transfer Dissociation through Fixed Charge Derivatization of Cysteines. *Anal. Chem.* **2009**, *81* (19), 7876–7884.

(74) Ko, B. J.; Brodbelt, J. S. Enhanced Electron Transfer Dissociation of Peptides Modified at C-Terminus with Fixed Charges. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (11), 1991–2000.

(75) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. Probing Protein Tertiary Structure with Amidination. *Anal. Chem.* **2005**, *77* (22), 7274–7281.

(76) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.

(77) Downard, K. M. Ions of the Interactome: The Role of MS in the Study of Protein Interactions in Proteomics and Structural Biology. *PROTEOMICS* **2006**, *6* (20), 5374–5384.

(78) Vasicek, L.; O'Brien, J. P.; Browning, K. S.; Tao, Z.; Liu, H.-W.; Brodbelt, J. S. Mapping Protein Surface Accessibility via an Electron Transfer Dissociation Selectively Cleavable Hydrazone Probe. *Mol. Cell. Proteomics MCP* **2012**, *11* (7), O111.015826.

(79) Mendoza, V. L.; Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* **2009**, *28* (5), 785–815.

(80) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–999.

(81) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K. A.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.

(82) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in Saccharomyces Cerevisiae Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154–1169.

(83) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the

Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(84) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(85) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(86) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(87) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(88) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(89) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(90) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(91) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(92) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(93) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(94) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(95) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

Chapter 2

(1) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. Anal. Chem. **2009**, 81 (9), 3645–3653.

(2) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. Anal. Chem. **2009**, 81 (19), 8109–8118.

Chapter 3

(1) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312* (5771), 212–217.

(2) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(3) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(4) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(5) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(6) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(7) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(8) Hughes, C.; Ma, B.; Lajoie, G. A. De Novo Sequencing Methods in Proteomics. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.

(9) Standing, K. G. Peptide and Protein de Novo Sequencing by Mass Spectrometry. *Curr. Opin. Struct. Biol.* **2003**, *13* (5), 595–601.

(10) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(11) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(12) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(13) Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (11), 2594–2604.

(14) Zhang, Z. De Novo Peptide Sequencing Based on a Divide-and-Conquer Algorithm and Peptide Tandem Spectrum Simulation. *Anal. Chem.* **2004**, *76* (21), 6374–6383.

(15) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(16) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(17) Bern, M.; Goldberg, D. De Novo Analysis of Peptide Tandem Mass Spectra by Spectral Graph Partitioning. *J. Comput. Biol.* **2006**, *13* (2), 364–378.

(18) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(19) Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q. pNovo: De Novo Peptide Sequencing and Identification Using HCD Spectra. *J. Proteome Res.* **2010**, *9* (5), 2713–2724.

(20) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(21) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(22) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(23) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(24) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(25) Grill, V.; Shen, J.; Evans, C.; Cooks, R. G. Collisions of Ions with Surfaces at Chemically Relevant Energies: Instrumentation and Phenomena. *Rev. Sci. Instrum.* **2001**, *72* (8), 3149–3179.

(26) Brodbelt, J. Shedding Light on the Frontier of Photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.

(27) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(28) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(29) Brodbelt, J. S.; Wilson, J. J. Infrared Multiphoton Dissociation in Quadrupole Ion Traps. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.

(30) Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Müller, R.; Böcker, S.; Huber, C. G.; Kohlbacher, O. De Novo Peptide Sequencing by Tandem MS Using Complementary CID and Electron Transfer Dissociation. *ELECTROPHORESIS* **2009**, *30* (21), 3736–3747.

(31) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(32) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.

(33) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(34) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(35) Smith, S. I.; Brodbelt, J. S. Characterization of Oligodeoxynucleotides and Modifications by 193 Nm Photodissociation and Electron Photodetachment Dissociation. *Anal. Chem.* **2010**, *82* (17), 7218–7226.

(36) Madsen, J. A.; Cullen, T. W.; Trent, M. S.; Brodbelt, J. S. IR and UV Photodissociation as Analytical Tools for Characterizing Lipid A Structures. *Anal. Chem.* **2011**, *83* (13), 5107–5113.

(37) Hankins, J. V.; Madsen, J. A.; Giles, D. K.; Brodbelt, J. S.; Trent, M. S. Amino Acid Addition to Vibrio Cholerae LPS Establishes a Link between Surface Remodeling in Gram-Positive and Gram-Negative Bacteria. *Proc. Natl. Acad. Sci.* **2012**, *109* (22), 8722–8727.

(38) Morgan, J. W.; Russell, D. H. Comparative Studies of 193-Nm Photodissociation and TOF-TOFMS Analysis of Bradykinin Analogues: The Effects of Charge Site(s) and Fragmentation Timescales. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (5), 721–729.

(39) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of Phosphorylation Site-Specific Dissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.

(40) Yoon, S. H.; Moon, J. H.; Kim, M. S. Dissociation Mechanisms and Implication for the Presence of Multiple Conformations for Peptide Ions with Arginine at the C-Terminus: Time-Resolved Photodissociation Study. *J. Mass Spectrom. JMS* **2010**, *45* (7), 806–814.

(41) Devakumar, A.; Mechref, Y.; Kang, P.; Novotny, M. V.; Reilly, J. P. Laser-Induced Photofragmentation of Neutral and Acidic Glycans inside an Ion-Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2007**, *21* (8), 1452–1460.

(42) Zhang, L.; Reilly, J. P. Peptide Photodissociation with 157 Nm Light in a Commercial Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2009**, *81* (18), 7829–7838.

(43) Thompson, M.; Cui, W.; Reilly, J. Factors That Impact the Vacuum Ultraviolet Photofragmentation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (8), 1439–1452.

(44) Diedrich, J. K.; Julian, R. R. Site Selective Fragmentation of Peptides and Proteins at Quinone Modified Cysteine Residues Investigated by ESI-MS. *Anal. Chem.* **2010**, *82* (10), 4006–4014.

(45) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(46) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(47) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. Tyrosine Sulfation in a Gram-Negative Bacterium. *Nat. Commun.* **2012**, *3*, 1153.

(48) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(49) Parthasarathi, R.; He, Y.; Reilly, J. P.; Raghavachari, K. New Insights into the Vacuum UV Photodissociation of Peptides. *J. Am. Chem. Soc.* **2010**, *132* (5), 1606–1610.

(50) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(51) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(52) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(53) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. *Anal. Chem.* **2005**, *77* (18), 5800–5813.

(54) Zhang, Z. Prediction of Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76* (14), 3908–3922.

(55) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(56) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(57) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(58) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo

Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(59) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(60) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(61) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(62) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(63) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(64) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(65) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(66) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(67) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(68) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(69) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(70) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(71) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(72) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(73) An, M.; Zou, X.; Wang, Q.; Zhao, X.; Wu, J.; Xu, L.-M.; Shen, H.-Y.; Xiao, X.; He, D.; Ji, J. High-Confidence de Novo Peptide Sequencing Using Positive Charge Derivatization and Tandem MS Spectra Merging. *Anal. Chem.* **2013**, *85* (9), 4530–4537.

(74) Der, B. S.; Kluwe, C.; Miklos, A. E.; Jacak, R.; Lyskov, S.; Gray, J. J.; Georgiou, G.; Ellington, A. D.; Kuhlman, B. Alternative Computational Protocols for Supercharging Protein Surfaces for Reversible Unfolding and Retention of Stability. *PLoS ONE* **2013**, *8* (5), e64363.

(75) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

Chapter 4

(1) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. De Novo Sequencing of Peptides by MS/MS. *PROTEOMICS* **2010**, *10* (4), 634–649.

(2) Ma, J.; Ward, C. C.; Jungreis, I.; Slavoff, S. A.; Schwaid, A. G.; Neveu, J.; Budnik, B. A.; Kellis, M.; Saghatelian, A. Discovery of Human sORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**.

(3) Noga, M. J.; Lewandowski, J. J.; Suder, P.; Silberring, J. An Enhanced Method for Peptides Sequencing by N-Terminal Derivatization and MS. *Proteomics* **2005**, *5* (17), 4367–4375.

(4) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(5) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(6) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(7) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(8) Brodbelt, J. Shedding Light on the Frontier of Photodissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.

(9) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(10) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(11) Taylor, J. A.; Johnson, R. S. Implementation and Uses of Automated de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (11), 2594–2604.

(12) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(13) Zhang, Z. De Novo Peptide Sequencing Based on a Divide-and-Conquer Algorithm and Peptide Tandem Spectrum Simulation. *Anal. Chem.* **2004**, *76* (21), 6374–6383.

(14) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(15) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. NovoHMM: A Hidden Markov Model for de Novo Peptide Sequencing. *Anal. Chem.* **2005**, *77* (22), 7265–7273.

(16) Bern, M.; Goldberg, D. De Novo Analysis of Peptide Tandem Mass Spectra by Spectral Graph Partitioning. *J. Comput. Biol.* **2006**, *13* (2), 364–378.

(17) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(18) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. A High-Throughput de Novo

Sequencing Approach for Shotgun Proteomics Using High-Resolution Tandem Mass Spectrometry. *BMC Bioinformatics* **2010**, *11* (1), 118.

(19) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(20) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(21) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(22) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(23) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(24) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(25) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(26) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(27) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(28) Beardsley, R.; Reilly, J. Fragmentation of Amidinated Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (2), 158–167.

(29) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(30) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(31) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(32) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(33) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(34) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(35) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(36) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(37) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(38) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.;

Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(39) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(40) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(41) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. Straightforward Ladder Sequencing of Peptides Using a Lys-N Metalloendopeptidase. *Nat. Methods* **2008**, *5* (5), 405–407.

(42) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation. *Anal. Chem.* **2009**, *81* (18), 7814–7822.

(43) Cannon, J. R.; Edwards, N. J.; Fenselau, C. Mass-Biased Partitioning to Enhance Middle down Proteomics Analysis. *J. Mass Spectrom.* **2013**, *48* (3), 340–343.

(44) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG

Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(45) Hansen, T. A.; Kryuchkov, F.; Kjeldsen, F. Reduction in Database Search Space by Utilization of Amino Acid Composition Information from Electron Transfer Dissociation and Higher-Energy Collisional Dissociation Mass Spectra. *Anal. Chem.* **2012**, *84* (15), 6638–6645.

(46) Angel, P. M.; Orlando, R. Quantitative Carbamylation as a Stable Isotopic Labeling Method for Comparative Proteomics. *Rapid Commun. Mass Spectrom.* **2007**, *21* (10), 1623–1634.

(47) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

(48) Jeong, K.; Kim, S.; Pevzner, P. A. UniNovo: A Universal Tool for de Novo Peptide Sequencing. *Bioinformatics* **2013**, *29* (16), 1953–1962.

Chapter 5

(1) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(2) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *ELECTROPHORESIS* **1999**, *20* (18), 3551–3567.

(3) Xu, H.; Freitas, M. A. MassMatrix: A Database Search Program for Rapid Characterization of Proteins and Peptides from Tandem Mass Spectrometry Data. *PROTEOMICS* **2009**, *9* (6), 1548–1555.

(4) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(5) Craig, R.; Cortens, J. P.; Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.

(6) De Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. Comprehensive Mass-Spectrometry-Based Proteome Quantification of Haploid versus Diploid Yeast. *Nature* **2008**, *455* (7217), 1251–1254.

(7) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(8) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.

(9) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: A Dynamic Programming Algorithm for de Novo Peptide Sequencing via Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(10) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973.

(11) Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q. pNovo: De Novo Peptide Sequencing and Identification Using HCD Spectra. *J. Proteome Res.* **2010**, *9* (5), 2713–2724.

(12) Jeong, K.; Kim, S.; Pevzner, P. A. UniNovo: A Universal Tool for de Novo Peptide Sequencing. *Bioinformatics* **2013**, *29* (16), 1953–1962.

(13) Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate de Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (24), 2348–2356.

(14) Keough, T.; Lacey, M. P.; Youngquist, R. S. Solid-phase Derivatization of Tryptic Peptides for Rapid Protein Identification by Matrix-assisted Laser Desorption/ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16* (11), 1003–1015.

(15) Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(16) Wang, D.; Kalb, S. R.; Cotter, R. J. Improved Procedures for N-Terminal Sulfonation of Peptides for Matrix-Assisted Laser Desorption/ionization Post-Source Decay Peptide Sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (1), 96–102.

(17) Lee, Y. H.; Han, H.; Chang, S.-B.; Lee, S.-W. Isotope-Coded N-Terminal Sulfonation of Peptides Allows Quantitative Proteomic Analysis with Increased de Novo Peptide Sequencing Capability. *Rapid Commun. Mass Spectrom.* **2004**, *18* (24), 3019–3027.

(18) Summerfield, S. G.; Bolgar, M. S.; Gaskell, S. J. Promotion and Stabilization of b1 Ions in Peptide Phenythiocarbamoyl Derivatives: Analogies with Condensed-Phase Chemistry. *J. Mass Spectrom.* **1997**, *32* (2), 225–231.

(19) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de Novo Sequencing Facilitated by a Dual-Labeling Strategy. *Anal. Chem.* **2005**, *77* (19), 6300–6309.

(20) Wilson, J. J.; Brodbelt, J. S. Infrared Multiphoton Dissociation for Enhanced de Novo Sequence Interpretation of N-Terminal Sulfonated Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (19), 6855–6862.

(21) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(22) Altelaar, A. F. M.; Navarro, D.; Boekhorst, J.; van Breukelen, B.; Snel, B.; Mohammed, S.; Heck, A. J. R. Database Independent Proteomics Analysis of the Ostrich and Human Proteome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (2), 407–412.

(23) Kim, J.-S.; Song, J.-S.; Kim, Y.; Park, S.; Kim, H.-J. De Novo Analysis of Protein N-Terminal Sequence Utilizing MALDI Signal Enhancing Derivatization with Br Signature. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1911–1919.

(24) Samgina, T. Y.; Kovalev, S. V.; Gorshkov, V. A.; Artemenko, K. A.; Poljakov, N. B.; Lebedev, A. T. N-Terminal Tagging Strategy for De Novo Sequencing of Short Peptides by ESI-MS/MS and MALDI-MS/MS. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 104–111.

(25) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and de Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteomics MCP* **2009**, *8* (4), 650–660.

(26) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O.; Tsunasawa, S. A Method for N-Terminal de Novo Sequencing of Nα-Blocked Proteins by Mass Spectrometry. *The Analyst* **2011**, *136* (1), 113.

(27) Hennrich, M. L.; Mohammed, S.; Altelaar, A. F. M.; Heck, A. J. R. Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965.

(28) Yamaguchi, M.; Oka, M.; Nishida, K.; Ishida, M.; Hamazaki, A.; Kuyama, H.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S.; Nishimura, O.; Tsunasawa, S.; Nakazawa, T. Enhancement of MALDI-MS Spectra of C-Terminal Peptides by the Modification of Proteins via an Active Ester Generated in Situ from an Oxazolone. *Anal. Chem.* **2006**, *78* (22), 7861–7869.

(29) Nakajima, C.; Kuyama, H.; Nakazawa, T.; Nishimura, O. C-Terminal Sequencing of Protein by MALDI Mass Spectrometry through the Specific Derivatization of the A-

Carboxyl Group with 3-Aminopropyltris-(2,4,6-Trimethoxyphenyl)phosphonium Bromide. *Anal. Bioanal. Chem.* **2012**, *404* (1), 125–132.

(30) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(31) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(32) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(33) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.

(34) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(35) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(36) Liu, X.; Shan, B.; Xin, L.; Ma, B. Better Score Function for Peptide Identification with ETD MS/MS Spectra. *BMC Bioinformatics* **2010**, *11* (Suppl 1), S4.

(37) Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009**, *81* (9), 3645–3653.

(38) Richards, A. L.; Vincent, C. E.; Guthals, A.; Rose, C. M.; Westphall, M. S.; Bandeira, N.; Coon, J. J. Neutron-Encoded Signatures Enable Product Ion Annotation from Tandem Mass Spectra. *Mol. Cell. Proteomics MCP* **2013**, *12* (12), 3812–3823.

(39) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(40) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.

(41) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Anal. Chem.* **2012**, *84* (5), 2433–2439.

(42) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(43) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(44) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. Proteomics-Grade de Novo Sequencing Approach. *J. Proteome Res.* **2005**, *4* (6), 2348–2354.

(45) Robotham, S. A.; Horton, A. P.; Cannon, J. R.; Cotham, V. C.; Marcotte, E. M.; Brodbelt, J. S. UVnovo: A Novel De Novo Sequencing Algorithm Using Single Series of Fragment Ions via Chromophore Tagging and 351 Nm UVPD. *Submitted*.

(46) Datta, R.; Bern, M. Spectrum Fusion: Using Multiple Mass Spectra for de Novo Peptide Sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **2009**, *16* (8), 1169–1182.

(47) An, M.; Zou, X.; Wang, Q.; Zhao, X.; Wu, J.; Xu, L.-M.; Shen, H.-Y.; Xiao, X.; He, D.; Ji, J. High-Confidence de Novo Peptide Sequencing Using Positive Charge Derivatization and Tandem MS Spectra Merging. *Anal. Chem.* **2013**, *85* (9), 4530–4537.

(48) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

Chapter 6

(1) Swaney, D. L.; McAlister, G. C.; Coon, J. J. Decision Tree–driven Tandem Mass Spectrometry for Shotgun Proteomics. *Nat. Methods* **2008**, *5* (11), 959–964.

(2) Bailey, D. J.; Rose, C. M.; McAlister, G. C.; Brumbaugh, J.; Yu, P.; Wenger, C. D.; Westphall, M. S.; Thomson, J. A.; Coon, J. J. Instant Spectral Assignment for Advanced Decision Tree-Driven Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2012**, *109* (22), 8411–8416.

(3) Pan, C.; Zhou, Y.; Dator, R.; Ginghina, C.; Zhao, Y.; Movius, J.; Peskind, E.; Zabetian, C. P.; Quinn, J.; Galasko, D.; Stewart, T.; Shi, M.; Zhang, J. Targeted Discovery and Validation of Plasma Biomarkers of Parkinson's Disease. *J. Proteome Res.* **2014**.

(4) Ma, J.; Ward, C. C.; Jungreis, I.; Slavoff, S. A.; Schwaid, A. G.; Neveu, J.; Budnik, B. A.; Kellis, M.; Saghatelian, A. Discovery of Human sORF-Encoded Polypeptides (SEPs) in Cell Lines and Tissue. *J. Proteome Res.* **2014**, *13* (3), 1757–1765.

(5) Seneviratne, U.; Godoy, L. C.; Wishnok, J. S.; Wogan, G. N.; Tannenbaum, S. R. Mechanism-Based Triarylphosphine-Ester Probes for Capture of Endogenous RSNOs. *J. Am. Chem. Soc.* **2013**, *135* (20), 7693–7704.

(6) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(7) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F. M.; Mechtler, K.; Heck, A. J. R.; Mohammed, S. Unambiguous Phosphosite Localization Using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *J. Proteome Res.* **2013**, *12* (3), 1520–1525.

(8) Mommen, G. P. M.; Frese, C. K.; Meiring, H. D.; Brink, J. van G. den; Jong, A. P. J. M. de; Els, C. A. C. M. van; Heck, A. J. R. Expanding the Detectable HLA Peptide Repertoire Using Electron-Transfer/higher-Energy Collision Dissociation (EThcD). *Proc. Natl. Acad. Sci.* **2014**, *111* (12), 4507–4512.

(9) Gunawardena, H. P.; He, M.; Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; Hodges, B. D. M.; McLuckey, S. A. Electron Transfer versus Proton Transfer in Gas-Phase Ion/Ion Reactions of Polyprotonated Peptides. *J. Am. Chem. Soc.* **2005**, *127* (36), 12627–12639.

(10) McGee, W. M.; McLuckey, S. A. The Ornithine Effect in Peptide Cation Dissociation: The Ornithine Effect: Selective Cleavages Observed in Ornithinated Peptides. *J. Mass Spectrom.* **2013**, *48* (7), 856–861.

(11) Krusemark, C. J.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. Modifying the Charge State Distribution of Proteins in Electrospray Ionization Mass Spectrometry by Chemical Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (9), 1617–1625.

(12) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (24), 9243–9252.

(13) Laskin, J.; Futrell, J. H. Collisional Activation of Peptide Ions in FT-ICR Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 158–181.

(14) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; A. L. Burlingame, Ed.; Academic Press, 2005; Vol. Volume 402, pp 148–185.

(15) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. Mass Spectrometry of Peptides and Proteins. *Methods* **2005**, *35* (3), 211–222.

(16) Beys-da-Silva, W. O.; Santi, L.; Berger, M.; Calzolari, D.; Passos, D. O.; Guimarães, J. A.; Moresco, J. J.; Yates, J. R. Secretome of the Biocontrol Agent Metarhizium Anisopliae Induced by the Cuticle of the Cotton Pest Dysdercus Peruvianus Reveals New Insights into Infection. *J. Proteome Res.* **2014**, *13* (5), 2282–2296.

(17) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1811–1822.

(18) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.

(19) Peng, Y.; Chen, X.; Zhang, H.; Xu, Q.; Hacker, T. A.; Ge, Y. Top-down Targeted Proteomics for Deep Sequencing of Tropomyosin Isoforms. *J. Proteome Res.* **2013**, *12* (1), 187–198.

(20) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *PROTEOMICS* **2008**, *8* (21), 4466–4483.

(21) Rose, C. M.; Russell, J. D.; Ledvina, A. R.; McAlister, G. C.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J.; Syka, J. E. P. Multipurpose Dissociation Cell for Enhanced ETD of Intact Protein Species. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (6), 816–827.

(22) Laskin, J.; Futrell, J. H. Surface-Induced Dissociation of Peptide Ions: Kinetics and Dynamics. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (12), 1340–1347.

(23) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-Induced Dissociation of Small Molecules, Peptides, and Non-Covalent Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 190–208.

(24) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.

(25) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.

(26) Chingin, K.; Makarov, A.; Denisov, E.; Rebrov, O.; Zubarev, R. A. Fragmentation of Positively-Charged Biological Ions Activated with a Beam of High-Energy Cations. *Anal. Chem.* **2014**, *86* (1), 372–379.

(27) Cook, S. L.; Collin, O. L.; Jackson, G. P. Metastable Atom-Activated Dissociation Mass Spectrometry: Leucine/isoleucine Differentiation and Ring Cleavage of Proline Residues. *J. Mass Spectrom.* **2009**, *44* (8), 1211–1223.

(28) Flora, J. W.; Muddiman, D. C. Determination of the Relative Energies of Activation for the Dissociation of Aromatic versus Aliphatic Phosphopeptides by ESI-FTICR-MS and IRMPD. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (1), 121–127.

(29) Brodbelt, J. S.; Wilson, J. J. Infrared Multiphoton Dissociation in Quadrupole Ion Traps. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.

(30) Vasicek, L. A.; Wilson, J. J.; Brodbelt, J. S. Improved Infrared Multiphoton Dissociation of Peptides through N-Terminal Phosphonite Derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384.

(31) Gardner, M. W.; Smith, S. I.; Ledvina, A. R.; Madsen, J. A.; Coon, J. J.; Schwartz, J. C.; Stafford, G. C.; Brodbelt, J. S. Infrared Multiphoton Dissociation of Peptide Cations in a Dual Pressure Linear Ion Trap Mass Spectrometer. *Anal. Chem.* **2009**, *81* (19), 8109–8118.

(32) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(33) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(34) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.

(35) Zhang, L.; Reilly, J. P. Peptide de Novo Sequencing Using 157 Nm Photodissociation in a Tandem Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2010**, *82* (3), 898–908.

(36) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.

(37) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein

Characterization Using Top-down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(38) Zhang, L.; Reilly, J. P. Radical-Driven Dissociation of Odd-Electron Peptide Radical Ions Produced in 157 Nm Photodissociation. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (7), 1378–1390.

(39) Parthasarathi, R.; He, Y.; Reilly, J. P.; Raghavachari, K. New Insights into the Vacuum UV Photodissociation of Peptides. *J. Am. Chem. Soc.* **2010**, *132* (5), 1606–1610.

(40) Cui, W.; Thompson, M. S.; Reilly, J. P. Pathways of Peptide Ion Fragmentation Induced by Vacuum Ultraviolet Light. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.

(41) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 Nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19* (12), 1657–1665.

(42) Wilson, J. J.; Brodbelt, J. S. Ultraviolet Photodissociation at 355 Nm of Fluorescently Labeled Oligosaccharides. *Anal. Chem.* **2008**, *80* (13), 5186–5196.

(43) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. De Novo Sequencing of Peptides Using Selective 351 Nm Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (20), 9832–9838.

(44) Cotham, V. C.; Wine, Y.; Brodbelt, J. S. Selective 351 Nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. *Anal. Chem.* **2013**, *85* (11), 5577–5585.

(45) Aponte, J. R.; Vasicek, L.; Swaminathan, J.; Xu, H.; Koag, M. C.; Lee, S.; Brodbelt, J. S. Streamlining Bottom-up Protein Identification Based on Selective Ultraviolet Photodissociation (UVPD) of Chromophore-Tagged Histidine- and Tyrosine-Containing Peptides. *Anal. Chem.* **2014**, *86* (13), 6237–6244.

(46) Joly, L.; Antoine, R.; Broyer, M.; Dugourd, P.; Lemoine, J. Specific UV Photodissociation of Tyrosyl-Containing Peptides in Multistage Mass Spectrometry. *J. Mass Spectrom.* **2007**, *42* (6), 818–824.

(47) Oh, J. Y.; Moon, J. H.; Kim, M. S. Sequence- and Site-Specific Photodissociation at 266 Nm of Protonated Synthetic Polypeptides Containing a Tryptophanyl Residue. *Rapid Commun. Mass Spectrom.* **2004**, *18* (22), 2706–2712.

(48) Oh, J. Y.; Moon, J. H.; Kim, M. S. Chromophore Effect in Photodissociation at 266 Nm of Protonated Peptides Generated by Matrix-Assisted Laser Desorption Ionization (MALDI). *J. Mass Spectrom.* **2005**, *40* (7), 899–907.

(49) Oh, J. Y.; Moon, J. H.; Lee, Y. H.; Hyung, S.-W.; Lee, S.-W.; Kim, M. S. Photodissociation Tandem Mass Spectrometry at 266 Nm of an Aliphatic Peptide Derivatized with Phenyl Isothiocyanate and 4-Sulfophenyl Isothiocyanate. *Rapid Commun. Mass Spectrom.* **2005**, *19* (10), 1283–1288.

(50) Park, S.; Ahn, W.-K.; Lee, S.; Han, S. Y.; Rhee, B. K.; Oh, H. B. Ultraviolet Photodissociation at 266 Nm of Phosphorylated Peptide Cations. *Rapid Commun. Mass Spectrom.* **2009**, *23* (23), 3609–3620.

(51) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6455–6458.

(52) Diedrich, J. K.; Julian, R. R. Site Selective Fragmentation of Peptides and Proteins at Quinone Modified Cysteine Residues Investigated by ESI-MS. *Anal. Chem.* **2010**, *82* (10), 4006–4014.

(53) Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358.

(54) Liu, Z.; Julian, R. R. Deciphering the Peptide Iodination Code: Influence on Subsequent Gas-Phase Radical Generation with Photodissociation ESI-MS. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (6), 965–971.

(55) Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609.

(56) Diedrich, J. K.; Julian, R. R. Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. *Anal. Chem.* **2011**, *83* (17), 6818–6826.

(57) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of Phosphorylation Site-Specific Dissociation of Singly Protonated Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.

(58) Sun, Q.; Yin, S.; Loo, J. A.; Julian, R. R. Radical Directed Dissociation for Facile Identification of Iodotyrosine Residues Using Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2010**, *82* (9), 3826–3833.

(59) Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.

(60) Enjalbert, Q.; Simon, R.; Salvador, A.; Antoine, R.; Redon, S.; Ayhan, M. M.; Darbour, F.; Chambert, S.; Bretonnière, Y.; Dugourd, P.; Lemoine, J. Photo-SRM: Laser-Induced Dissociation Improves Detection Selectivity of Selected Reaction Monitoring Mode. *Rapid Commun. Mass Spectrom.* **2011**, *25* (22), 3375–3381.

(61) Enjalbert, Q.; Girod, M.; Simon, R.; Jeudy, J.; Chirot, F.; Salvador, A.; Antoine, R.; Dugourd, P.; Lemoine, J. Improved Detection Specificity for Plasma Proteins by Targeting Cysteine-Containing Peptides with Photo-SRM. *Anal. Bioanal. Chem.* **2013**, *405* (7), 2321–2331.

(62) Girod, M.; Biarc, J.; Enjalbert, Q.; Salvador, A.; Antoine, R.; Dugourd, P.; Lemoine, J. Implementing Visible 473 Nm Photodissociation in a Q-Exactive Mass Spectrometer: Towards Specific Detection of Cysteine-Containing Peptides. *Analyst* **2014**, *139* (21), 5523–5530.

(63) Tecklenburg, R. E.; Miller, M. N.; Russell, D. H. Laser Ion Beam Photodissociation Studies of Model Amino Acids and Peptides. *J. Am. Chem. Soc.* **1989**, *111* (4), 1161–1171.

(64) Solouki, T.; Russell, D. H. Structural Mass Spectrometry of Matrix-Assisted Laser-Desorbed Biomolecules by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Photoionization and Photofragmentation. *Appl. Spectrosc.* **1993**, *47* (2), 211–217.

(65) Wilson, J. J.; Brodbelt, J. S. MS/MS Simplification by 355 Nm Ultraviolet Photodissociation of Chromophore-Derivatized Peptides in a Quadrupole Ion Trap. *Anal. Chem.* **2007**, *79* (20), 7883–7892.

(66) Lai, C.-K.; Ng, D. C. M.; Pang, H. F.; Le Blanc, J. C. Y.; Hager, J. W.; Fang, D.-C.; Cheung, A. S.-C.; Chu, I. K. Laser-Induced Dissociation of Singly Protonated Peptides at 193 and 266 Nm within a Hybrid Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2013**, *27* (10), 1119–1127.

(67) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom. JMS* **2000**, *35* (12), 1399–1406.

(68) Hu, Y.; Hadas, B.; Davidovitz, M.; Balta, B.; Lifshitz, C. Does IVR Take Place Prior to Peptide Ion Dissociation? *J. Phys. Chem. A* **2003**, *107* (34), 6507–6514.

VITA

Scott Allen Robotham was born in Lincoln, Nebraska. After graduating from Malcolm High School in the spring of 2004, he continued his education at Nebraska Wesleyan University in the fall of 2004. He graduated with a bachelor's of science degree in chemistry May 2009. Following graduation, he moved to Austin, TX to begin his graduate studies at the University of Texas at Austin under the direction of Professor Jennifer S. Brodbelt. His current work focuses on the development of mass spectrometric-based for the enhancement of data analysis algorithms for proteomic applications.

Permanent email: sarobotham@gmail.com

This dissertation was typed by the author.